目　录

第二部分 随机模型的建立

第三部份　传递函数模型的建立

第五部份 图 表

第六部分 习题和问题

1 引言

时间序列是按照时间顺序取得的一系列观测值．很多数据是以时间序列的形式出现的：一个工厂装船货物数量的月度序列，公路事故数量的周度序列，某化工过程产出的按小时观测，等等．时间序列的例子在一些领域中是极丰富的，诸如经济，商业，工程，自然科学 (特别是地球物理学和气象学) 和社会科学．第 4 章一开始的图 4.1 给出了时间序列的图形，这些都是该类数据的例子，也是我们将要讨论和涉及的．时间序列典型的一个本质特征就是相邻观测值的依赖性．时间序列观测值之间的这种依赖特征具有很大的实际意义．时间序列分析所论及的就是对这种依赖性进行分析的技巧．这要求对时间序列数据生成随机动态模型，并将这种模型用于重要的应用领域．

在本书后面的各章中，我们将给出对于时间序列和动态系统进行模型的建立、识别、拟合及检验的方法．所讨论的这些方法对于离散 (数据采样) 系统是适用的，即系统的观测发生于时间的相等间隔．

我们通过四个重要应用领域来说明时间序列和动态模型的应用．

1. 由时间序列的当前和过去值对未来值的预报．

2. 确定惯性系统的传递函数，即确定动态输入 - 输出模型，由此可描述任一给定输入序列对于系统输出的影响．

3. 在传递函数模型中使用示性输入变量来描述和评价异常干

预事件对时间序列特性的影响.

4. 设计简单的控制方案，由此可以通过对输入序列值的调整来尽可能地补偿系统输出对于理想目标值的潜在偏差。

1.1 四个重要的实际问题

1.1.1 时间序列的预报

利用一个时间序列在 t 时刻的有效观测值去预报在某个未来时刻 $t+l$ 该序列的值，这一问题的提出是基于下列事实：(1) 经济和商业计划，(2) 生产计划，(3) 库存和生产控制，以及 (4) 生产过程的控制和优化。正如最初 Holt 等等 [113],Brown[64], 以及帝国化工产业短期预报专题论文集中所述，预报往往要提前一段时间作出，即所谓的提前期 (lead time)，它是随着具体问题而不同。例如， Harrison[105] 将库存问题中的提前期定义为：从工厂做出补充库存订单开始，直到订货交付库存为止的一段时间.

我们假定观测值是在时间的离散、等间隔区间上得到的. 例如，在一个销售预测问题中，当前月份 t 的销售额 z_t 和已往月份的销售额 $z_{t-1}, z_{t-2}, z_{t-3}\cdots$ 可以用来预报提前 $l=1,2,3,\cdots,12$ 个月的销售额. 用 $\hat{z}_t(l)$ 记在原点 t 对未来某个时刻 $t+l$ 的销售额 z_{t+l} 所做的预报值，也就是提前期为 l 的预报. 函数 $\hat{z}_t(l), l=1,2,\cdots,$ 给出了在原点 t 对所有未来提前期的预报值，它被称为在原点 t 的预报函数. 我们的目的是得到一个这样的预报函数，使得对于每一提前期 l, 实际值与预报值之间偏差 $z_{t+l}-\hat{z}_t(l)$ 的均方尽可能地小.

除了计算最好的预报值之外，确定预报的精度也是必要的，由此，例如就可以把根据这些预报值所作决策的风险计算出来. 预报值的精度可以通过计算每个预报值两侧的概率限来表示. 这些概率限可以对任一组方便的概率算出，比如 50%,95% . 它们所指的是：最终发生的时间序列实现值将会以给定的概率落在这些

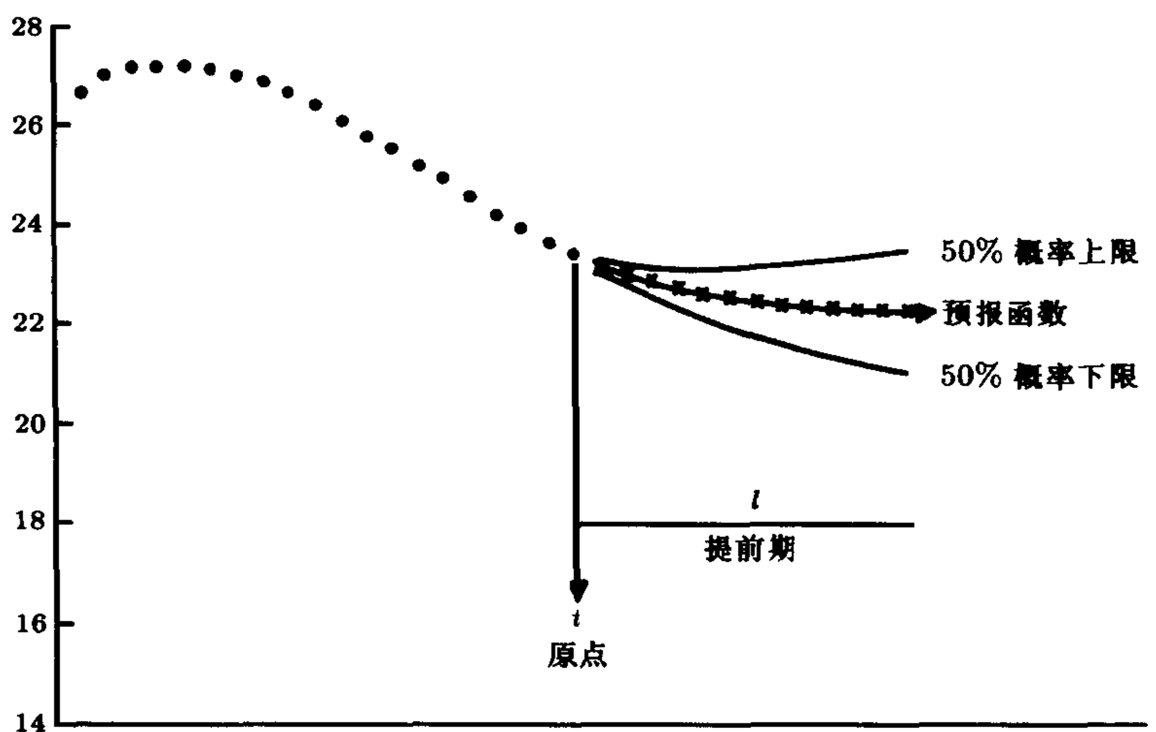


图 1.1 时间序列值与预报函数及 50% 概率限

界限内. 为了说明起见, 图 1.1 给出了以时刻 t 为终点的一个时间序列的最后 20 个值, 并给出在原点 t 所作的提前期 $l=1,2,\cdots,13$ 的预报值及 50% 的概率限。

得到预报值和估计概率限的方法将在第 5 章中做详细地讨论。推导这些预报方法是基于假定时间序列 z_t 遵从一个形式已知的随机模型。在第 3 章和第 4 章中引入的就是一类很有用的模型，这类模型适合于描述序列 z_t 的特性，称之为自回归求和滑动平均 (ARIMA) 模型，我们要研究这种模型的诸多性质。紧接着，在第 6,7,8 章中我们研究了对于实际的时间序列数据拟合这种模型的具体事项，相应的方法可通过三个步骤来说明：试探性模型识别或鉴别，模型参数的估计，模型的检验和诊断。

1.1.2 传递函数的估计

工业中一个很有意义的课题是研究过程的动态特性 [15],[115]。进行这种研究可使 (1) 实现对现有装置的更好控制，(2) 改进新

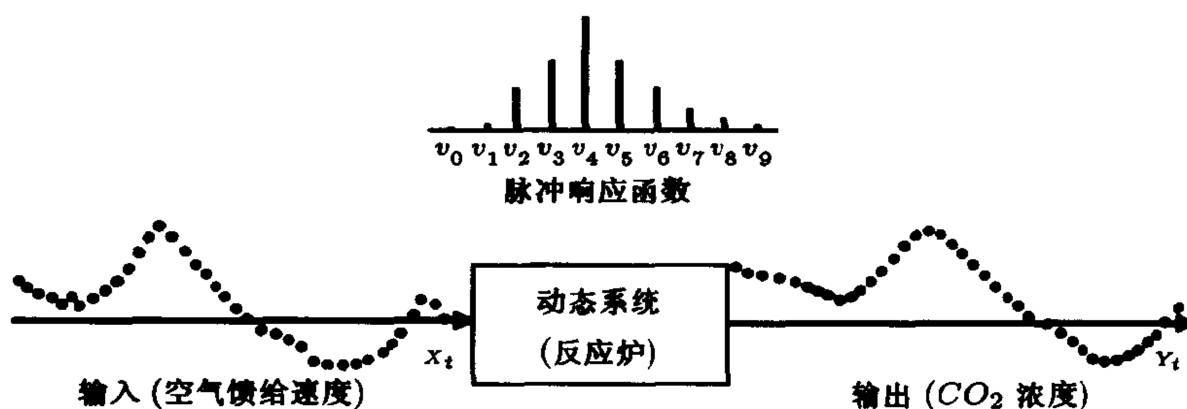


图 1.2　与一个动态系统有关的输入和输出时间序列

装置的设计. 具体来说, 已提出若干种方法, 都是从由输入时间序列 X_t 和输出时间序列 Y_t 所组成的过程的记录出发, 来估计设备单元的传递函数. 在图 1.2 中给出了一段这样的记录, 其中输入是空气供应的速度, 输出是反应炉中产生二氧化碳的浓度. 观测值以 9 秒为间隔得到. 图中还用棒形图的形式给出了一个假设的脉冲响应函数 $v_j, j = 1, 2, \cdots$, 它通过输入 X_t 与输出 Y_t 之间形如 $Y_t = \sum_{j=0}^{\infty} v_j Y_{t-j}$ 的动态线性关系决定了系统的传递函数. 第 10 章介绍了将输入过程 X_t 和输出过程 Y_t 联系起来的传递函数模型, 并考察了它们的许多性质. 根据确定性的扰动输入, 比如阶跃、脉冲和正弦等变化来估计传递函数的方法并不总是成功的. 这是因为, 对于大小适度、可以容忍的干扰来说, 系统的响应可能被不可控的干扰所屏蔽, 这些干扰统称为噪声. 估计传递函数的统计方法允许系统中有噪声, 在第 11 章中将讲述这种方法. 在经济学, 工程学, 生物学和很多其它领域中对于动态响应的估计都是非常有意义的.

传递函数模型的另一个重要应用是预报. 例如, 若能够确定出两个时间序列 X_t 和 Y_t 之间的动态关系, 则可以用两个序列的过去值对 Y_t 作预报. 在某些场合, 这种方法可以使预报误差大大降低.

1.1.3 异常干预事件对系统影响的分析

在一些场合，我们可以得知某些非常的外部事件、干预事件会影响被研究的时间序列 z_t 。这些干预事件的例子如：新环境法规的纳入，经济政策的变化，罢工以及特别的促销活动等。在这些情况下，正如在节 1.1.2 中所讨论的，我们可以用传递函数模型来解释干预事件对序列 z_t 的影响，但这里的"输入"序列将是简单的示性变量形式，即仅取值 1 或 0, 以表示 (定性地) 该事件出现或不出现。

在上述情形，干预分析用来获得干预事件对所关心时间序列影响的定量度量。作为例子， Box 和 Tiao[59] 利用干预模型进行数量化研究，他们研究了洛杉矶地区烟雾生成的氧化物对于空气污染控制的影响，以及美国消费价格指数对于经济控制的影响。另外，干预分析还可用于序列 z_t 中任意异常值的修正，这些异常值可能是由于干预事件引起的，这样就可以保证对序列进行时间序列分析的结果，例如拟合模型的结构，模型参数的估计以及未来值的预报都不致因异常值的影而被严重扭曲。在第 12 章中给出了干预分析模型及其应用，同时还研究了时间序列中野值或异常值检测的有关课题。

1.1.4 离散控制系统

在过去，对于统计学家来说，"过程控制"一词通常是指最初由 Shewhart[177](又见 Dudding 和 Jennet[85]) 在美国提出的质量控制技术。最近，有关质量控制的一系列方向受到了重视，由此导致 Page[154], [155] 及 Barnard[21] 的"累计求和图"(cumulative sum charts 和 Roberts[170] "几何滑动平均图"(geometric moving average charts) 的引入。这些图在非连续地制造"部件"的工业生产中经常使用，是所谓过程统计控制 (SPC) 的一个方向。具体地说 (见 Deming [79]), 它们被用于过程的连续监控，也就是说，用来支持一个连续监控的装置，以便检测发生变化的指定 (或具

体) 原因. 设备上的数据及时地显示, 以保证运行过程的重大变化能很快地被注意到. 知道了问题: "这种特定的变化是何时发生? " 的答案, 我们就可以去回答问题: "它为什么会发生? ", 于是对于过程连续施加激励手段, 以实现过程的稳定和改善.

与之对照, 在过程和化学工业中, 各种形式的反馈和前馈调节被用于我们称之为工程过程控制 (EPC) 中. 因为工程过程控制所作的调节通常是自动计算和实施的, 所以这类控制有时也称为自动化过程控制 (APC) . 但是, 实现这类调节的方式是方便的. 当在系统的输入中存在固有的干扰或噪声, 且不可能或无法将它消除, 这时这类控制就是必要的. 当我们测量到某个输入变量的波动时 — 可以观测但不能改变, 对另外的控制变量进行适当的补偿性修改, 则是可行的. 这就是前馈控制. 另外, 我们可以利

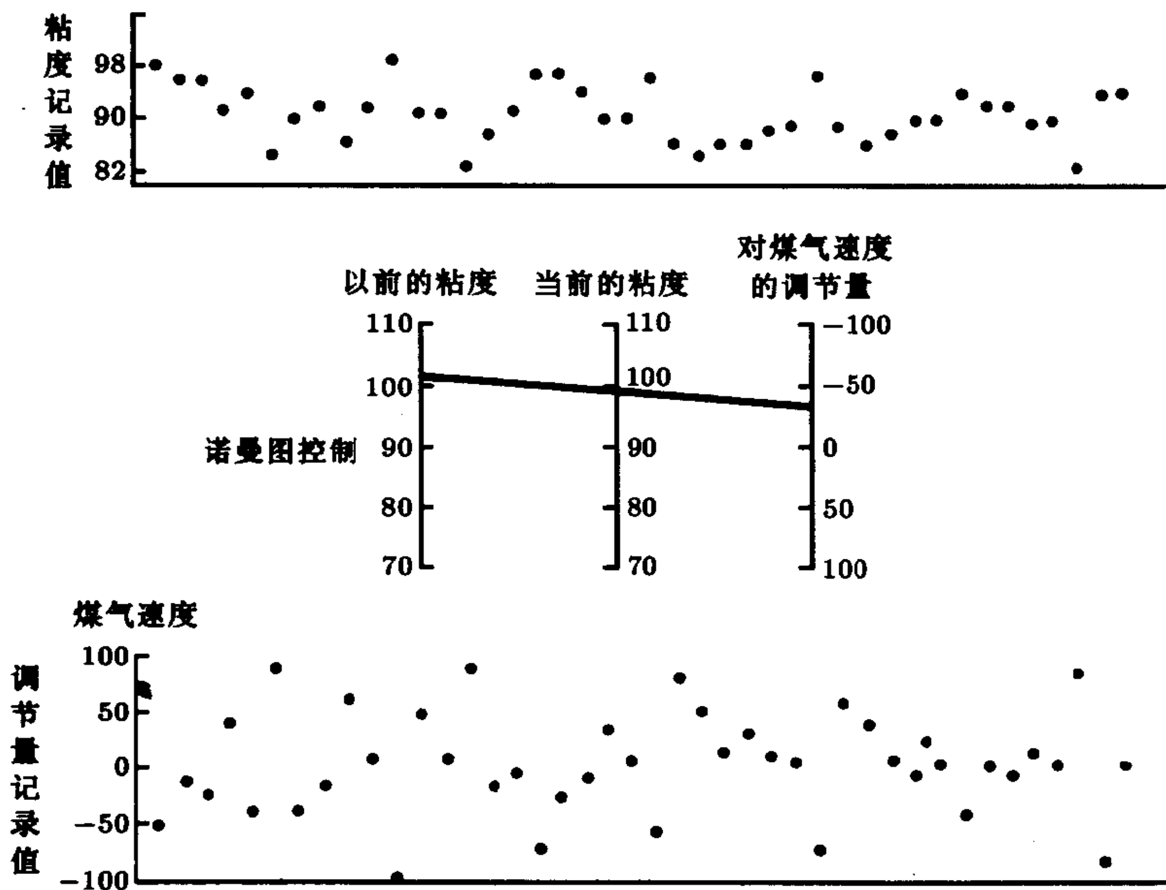


图 1.3　粘度的控制. 粘度观测的记录值和用诺曼图对煤气比例的调节值

用对于目标值的偏离或“错误信号”这类输出量本身的特征，由此计算出控制变量的补偿性改变值，这称为反馈控制。与前馈控制不同，即便在干扰源不完全知道或干扰的大小无法测出时，这种修正方式也能使用。

在第 13 章中，我们引用本书前面讨论的有关时间序列和传递函数模型的内容，来给出这些控制方法的统计解释，以便评价这些方法之间的关系及其不同的目的。具体地说，我们要说明：如何利用反馈控制的思想来设计人工调节生产过程简单的操作图。例如图 1.3 的上部给出了 42 小时的时段上聚合物粘度的每小时测量值。该粘度被控制在 90 个单位的目标值附近。当每个粘度测量值得到后，过程操作者就利用图中间给出的诺曼图 (nomogram) 计算出对操作变量 (煤气速度) 应作的调节值。图 1.3 的下部给出了按照诺曼图作出的调节值。

1.2 随机性和确定性的动态数学模型

用数学模型去描述一种物理现象的特性，这一思想是确信无疑的了。具体地，有时可以根据物理定律推出一个模型，使得我们对某个依赖于时间的量在任何瞬时都能几乎准确地计算其值。比如，我们可以计算沿已知方向、以已知速度发射的导弹的轨道。如果精确的计算是可能的，那么，这样的模型就是完全确定性的。

然而，或许没有一种现象是完全确定性的，因为可能出现一些未知的因素，比如，变化的风速能使导弹稍稍偏离轨线。在许多问题中，我们必须去考虑一种依赖于时间的现象 (例如，新闻纸的月销售量)，其中有很多未知因素，且不可能写出一个确定性的模型去计算该现象的未来特性。但是，推导出一个模型，能够用来计算某未来变量落在两个特定界限之间的概率，这是可能的。这样的模型称为概率模型或随机模型。时间序列所需要的模型 (如为了实现最优预报和控制) 实际上就是随机模型。接下来，我们必须将概率模型 (有时也称为随机模型) 和观测到的时间序列加

以区分. N 个顺次观测值 $z_1, z_2, \cdots, z_N$ 的时间序列被看作是随机过程所生成时间序列无穷总体中的一个样本实现. 我们常常把随机这个词从"随机过程"中省去, 而只称"过程".

1.2.1 用于预报和控制的平稳和非平稳随机模型

描述时间序列的一类重要随机模型受到了广泛的注意, 这就是所谓的平稳模型, 该类模型假设过程在一个不变的均值水平附近保持均衡. 但是, 在预报具有特别重要作用的工业、商业和经济中, 许多时间序列更适合用非平稳模型描述, 特别, 它们都不具有在时间上不变的均值水平. 因此, 并不令人感到惊讶, 由 Holt[112],[113],Winters[206],Brown[64] 最初提出的许多经济预报方法, 以及 ICI 专题论文 [178] 所使用的指数加权滑动平均等方法, 都可以说明是适合于某种特定类型的非平稳过程. 尽管这些方法过于狭窄, 不能有效地处理所有的时间序列, 但是它们给出了正确形式的预报函数, 这一事实提供了线索, 促使我们去寻求能对这些问题有用的一类非平稳模型.

指数滑动平均给出了最小均方误差预报 (Muth[147]), 所对应的随机模型是一类非平稳过程中的一员, 称为自回归求和滑动平均过程 (ARIMA 过程), 我们在第 4 章中将加以讨论. 这类过程提供了较广的模型范围: 平稳模型, 非平稳模型, 它们适合于描述实际中的许多时间序列. 正如第 5 章所述, 一旦对序列确定出合适的模型, 立刻就有最优预报方法. 指数滑动平均预报只是作为特例包含在这些预报方法中.

一些简单的算子 我们将广泛地使用后移算子 B,它定义为 $Bz_t = z_{t-1}$; 从而 $B^m z_t = z_{t-m}$. 其逆运算由前移算子 $F = B^{-1}$ 来实现, 是由 $Fz_t = z_{t+1}$ 给出; 从而 $F^m z_t = z_{t+m}$. 另一个重要的算子是向后差分算子 ∇, 它可用 B 表示, 因为

$$\nabla z_t = z_t - z_{t-1} = (1 - B) z_t$$

线性滤波器模型　我们所使用的随机模型都基于这样的思想(Yule[212])：顺序值之间高度依赖的时间序列常常可以看作是由一个独立的"冲击"序列所生成，这些"冲击"是由一个固定分布随机得出，且往往假定分布为正态，并具有零均值和方差 σ_a^2。这样的一串随机变量 $a_t, a_{t-1}, a_{t-2}, \cdots$ 称为白噪声过程。

我们假定，如图 1.4 所示，白噪声过程通过一个所谓的线性滤波器可以转化成过程 z_t。线性滤波运算可简单地取为既往随机冲击 a_t 的加权和，即

$$\begin{aligned} z_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= \mu + \psi(B) a_t \end{aligned} \tag{1.2.1}$$

通常，μ 是确定过程"水平"的参数，而

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$$

是将 a_t 转化为 z_t 的算子，称为滤波器的传递函数。

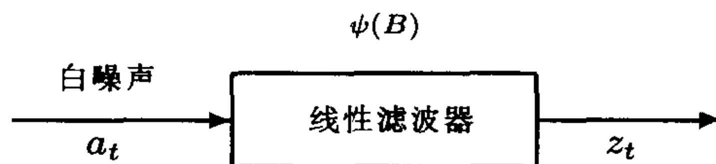


图 1.4　时间序列表示为线性滤波器的输出

在理论上，由权数构成的序列 $\psi_1, \psi_2, \cdots$ 可以是有限的，也可以是无限的。如过该序列是有限的，或者虽是无限的但为绝对可和即 $\sum_{j=0}^{\infty} |\psi_j| < \infty$，则滤波器被称为是稳定的，且过程 z_t 是平稳的。那么，参数 μ 就是该过程上下变动的平均值。反之，若 z_t 是非平稳的，μ 就没有特别的意义，只不过是过程水平的一个参照点而已。

自回归模型　在描述某些实际中出现的序列时，一种非常有用的随机模型就是自回归模型。在该模型中，过程的当前值被表示为

过程的有穷线性组合再加上一个冲击 a_t。我们用 $z_t, z_{t-1}, z_{t-2}, \cdots$ 记在等间隔时间 $t, t-1, t-2, \cdots$ 上的过程值. 另外, 用 $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \cdots$ 记关于均值 μ 的偏差; 例如 $\tilde{z}_t = z_t - \mu$。那么

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{1.2.2}$$

便叫做为 p 阶自回归 (AR) 过程. 称其为自回归是因为线性模型

$$\tilde{z} = \phi_1 \tilde{x}_1 + \phi_2 \tilde{x}_2 + \cdots + \phi_p \tilde{x}_p + a$$

把一个因变量 z 与一组自变量 $x_1, x_2, \cdots, x_p$ 联系在一起, 并加上一个误差项 a, 通常称之为回归模型, 并称 z 是关于 $x_1, x_2, \cdots, x_p$ 的回归. 在 (1.2.2) 中, 变量 z 是关于其自身既往值的回归, 因此, 就是自回归模型. 如果我们定义 p 阶自回归算子为

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

则自回归模型就可以简记为

$$\phi(B)\tilde{z}_t = a_t$$

该模型包含 $p+2$ 个未知参数 $\mu, \phi_1, \phi_2, \cdots, \phi_p, \sigma_a^2$, 在实际中这些参数必须由数据来估计. 附加参数 σ_a^2 是白噪声过程的方差.

不难看出, 自回归模型是线性滤波器模型 (1.2.1) 的特殊情形. 例如, 我们可以将

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \phi_2 \tilde{z}_{t-3} + \cdots + \phi_p \tilde{z}_{t-p-1} + a_{t-1}$$

代入 (1.2.2) 式的右端以消去 $\tilde{z}_{t-1}$. 类似地, 我们可以代换 $\tilde{z}_{t-2}$ 等等, 最后得到一个 a 的无穷级数. 用符号来表示, 我们有

$$\phi(B)\tilde{z}_t = a_t$$

等价于

$$\tilde{z}_t = \psi(B) a_t$$

其中

$$\psi(B) = \phi^{-1}(B)$$

自回归过程可能是平稳的或是非平稳的。为了使过程是平稳的，必须选取 ϕ 使得在 $\psi(B) = \phi^{-1}(B)$ 中的权数 $\psi_1, \psi_2, \cdots$ 构成一个收敛级数。平稳的必要条件是：自回归算子 $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ 被视为 B 的 p 次多项式时， $\phi(B) = 0$ 所有根的绝对值都必须大于 1; 也就是所有根都在单位圆外。

滑动平均模型 自回归模型 (1.2.2) 把过程的偏差 $\tilde{z}_t$ 表示为 p 个过去偏差 $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \cdots, \tilde{z}_{t-p}$ 的有限加权和，再加一个随机冲击 a_t。正如我们刚才所见，等价地可把 $\tilde{z}_t$ 表示为 a 的无限加权和。

在描述观测到的时间序列时，另一类在实际中很重要的模型是有限滑动平均模型。这里，我们使 $\tilde{z}_t$ 线性依赖于有限的 q 个 a 的过去值，于是

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{1.2.3}$$

称为 q 阶滑动平均 (MA) 过程。“滑动平均”一词可能会产生误解，因为事实上，乘在 a 上的权数 $1, -\theta_1, -\theta_2, \cdots, -\theta_q$ 不必总和为 1, 也不必是正的。

如果我们定义一个 q 阶滑动平均算子为

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

则滑动平均模型可简记为

$$\tilde{z}_t = \theta(B) a_t$$

其中包含 $q+2$ 个参数 $\mu, \theta_1, \cdots, \theta_q, \sigma_a^2$, 在实际中必须由数据来估计。

自回归滑动平均混合模型 为了在实际时间序列的拟合中能有较大的灵活性，有时将自回归和滑动平均项一同纳入模型是很有好处的。这就引出了自回归滑动平均混合模型

$$\widetilde{z}_t = \phi_1\widetilde{z}_{t-1} + \cdots + \phi_p\widetilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (1.2.4)$$

或

$$\phi(B)\widetilde{z}_t = \theta(B)a_t$$

其中用到 $p+q+2$ 个未知参数 $\mu;\phi_1,\cdots,\phi_p;\theta_1,\cdots,\theta_q;\sigma_a^2$，要由数据来估计。这个模型还可用线性滤波器 (1.2.1) 写成 $\widetilde{z}_t = \phi^{-1}(B)\theta(B)a_t$ 。事实上，为了对实际中发生的平稳时间序列作恰当的描述，往往能够得到自回归，滑动平均，或混合的模型，其阶数通常不超过 2, 常常是小于 2.

非平稳模型 在工业或商业中 (例如，股票价格) 实际遇到的许多序列表现出非平稳的特性，并不是围绕一个固定的均值而变化。然而这样的序列可能表现出某种相同的特征。具体说来，尽管波动起伏所围绕的一般水平在不同的时刻可能是不同的，但是在允许水平有差异的前提下，这些序列的广义特征可能是相似的。我们在第 4 章和其后的各章中说明了：这样的特征常常可以用一个广义的自回归算子 $\varphi(B)$ 来描述，其中多项式 $\varphi(B)$ 的一个或多个零点 (即方程 $\varphi(B)=0$ 的一个或多个根) 在单位圆上。特别地，若存在 d 个单位根，则算子 $\varphi(B)$ 可写为

$$\varphi(B) = \phi(B)(1-B)^d$$

其中 $\phi(B)$ 是平稳算子。这样，能够描述同质非平稳特征的模型具有如下形式

$$\varphi(B)z_t = \phi(B)(1-B)^d z_t = \theta(B)z_t$$

即

$$\phi(B)w_t = \theta(B)a_t \quad (1.2.5)$$

这里

$$w_t = \nabla^d z_t \tag{1.2.6}$$

因此，同质非平稳特征有时可以用一个平稳模型来描述，当然必须要求过程的 d 阶差分是平稳的。实际上，d 通常是 0,1, 或者最多是 2.

由 (1.2.5) 和 (1.2.6) 所定义的过程提供了描述平稳或非平稳时间序列的有效模型，称为 (p, d, q) 阶自回归求和滑动平均过程 (ARIMA)。该过程定义为

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{1.2.7}$$

这里，$w_t = \nabla^d z_t$。注意，如果在 $d = 0$ 时用 $z_t - \mu$ 代替 w_t, 则模型 (1.2.7) 包括平稳的混合模型 (1.2.4), 作为特例，也包括了纯自回归模型 (1.2.2) 和纯滑动平均模型 (1.2.3)。

在 ARIMA 这个名称中包含"求和"(integrated) 一词 (可能称为"相加"更合适), 其原因是：(1.2.6) 的逆转关系是 $z_t = S^d w_t$, 这里 $S = \nabla^{-1} = (1 - B)^{-1}$, 即是下式定义的求和算子:

$$Sw_t = \sum_{j=0}^{\infty} w_{t-j} = w_t + w_{t-1} + w_{t-2} + \cdots$$

因此，一般的求和自回归滑动平均 (ARIMA) 过程就是对平稳过程 w_t 作 d 次相加或求和而生成的。在第 9 章中我们叙述了如何用模型 (1.2.7) 的具体形式来描述季节时间序列。

1.2.2 传递函数模型

在许多物理实例中都可以发现连续的输入与输出之间存在动态关系，其中重要的一类是通过线性差分方程把输入 X 和输出 Y 关于适当均值的偏差联系起来。在第 10 章中，我们给出了在等时间观测的输入 X 和输出 Y 之间的传递关系，是用如下的差分方程表示:

$$(1 + \xi_1 \nabla + \cdots + \xi_r \nabla^r) Y_t = (\eta_0 + \eta_1 \nabla + \cdots + \eta_s \nabla^s) X_{t-b} \tag{1.2.8}$$

其中用差分算子 ∇ 替换了微分算子 $D = d/dt$。形如 (1.2.8) 的表达式只含有很少的参数 ($r < 2, s < 2$), 我们常常可用它来近似性质更复杂的真实动态系统。

在 (1.2.8) 中用 $B = 1 - \nabla$ 代入，可以将线性模型 (1.2.8) 等价地写成用输入和输出的过去值表示的形式，即

$$\begin{aligned}(1 - \delta_1 B - \cdots - \delta_r B^r)Y_t &= (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)X_{t-b} \\ &= (\omega_0 B^b - \omega_1 B^{b+1} - \cdots - \omega_s B^{b+s})X_t \end{aligned} \tag{1.2.9}$$

或

$$\begin{aligned}\delta(B)Y_t &= \omega(B)B^b X_t \\ &= \Omega(B)X_t\end{aligned}$$

另外，我们还可以说输出 Y_t 和输入 X_t 是通过一个滤波器相联系：

$$\begin{aligned}Y_t &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots \\ &= v(B)X_t\end{aligned} \tag{1.2.10}$$

其中的传递函数

$$v(B) = v_0 + v_1 B + v_2 B^2 + \cdots \tag{1.2.11}$$

可以表示为两个多项式之比;

$$v(B) = \frac{\Omega(B)}{\delta(B)} = \delta^{-1}(B)\Omega(B)$$

如果级数 (1.2.11) 在 $|B| < 1$ 时收敛，则称线性滤波器 (1.2.10) 是稳定的。传递函数 (1.2.11) 中出现的权序列 $v_0, v_1, v_2, \cdots$ 称为脉冲响应函数。我们注意到对于模型 (1.2.9) 来说，前 b 个权数 $v_0, v_1, \cdots, v_{b-1}$ 为零。图 1.2 的中间给出了该图所示系统假设的脉冲响应函数。

附加噪声的模型 我们已经看到，估计出一个适当的模型以便把输出 Y_t 和输入 X_t 联系起来，这等价于去估计传递函数 $v(B)=\delta^{-1}(B)\Omega(B)$ 。但是，在实际中这个问题由于噪声 N_t 的存在而变得复杂了。我们假设对于输出和输入之间真正关系的干扰是按照如下方式

$$Y_t = v(B)X_t + N_t$$

这里， N_t 和 X_t 是独立的。假设如图 1.5 所示，噪声 N_t 可以用一个形如 (1.2.5) 或 (1.2.7) 的非平稳随机模型来描述，即

$$N_t = \psi(B)a_t = \varphi^{-1}(B)\theta(B)a_t$$

那么观测到的输出与输入之间的关系就是

$$\begin{aligned} Y_t &= v(B)X_t + \psi(B)a_t \\ &= \delta^{-1}(B)\Omega(B)X_t + \varphi^{-1}(B)\theta(B)a_t \end{aligned} \tag{1.2.12}$$

在实际中，为了描述输入和输出之间的动态关系，要估计传递函数 $v(B)=\delta^{-1}(B)\Omega(B)$，此外，还必须估计描述噪声的线性滤波器相应的传递函数 $\psi(B)=\varphi^{-1}(B)\theta(B)$，在第 11 章中将讨论有关的实现方法。

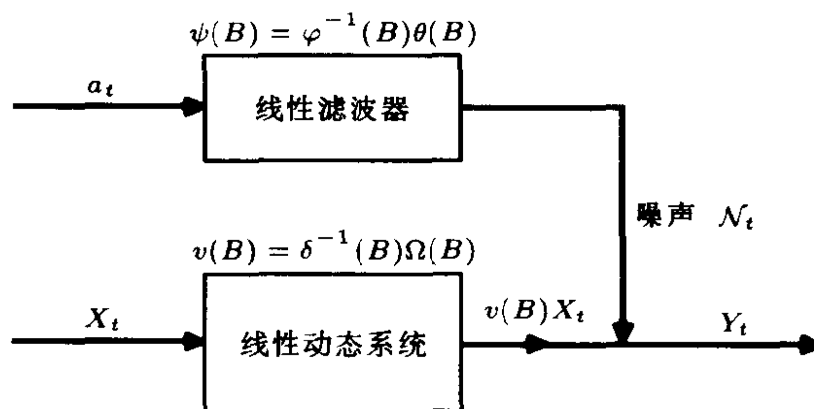


图 1.5 具有附加噪声模型的动态系统的传递函数模型

1.2.3 离散控制系统的模型

正如在节 1.1.4 中所述，控制就是对影响系统的干扰进行补偿的一种企图。这些干扰中有一些是可观测到的；另一些是不可观测的，而且只是作为对被控制特征目标值无法解释的偏离而出现。为了说明所涉及到的一般原理，考虑不可观测的干扰影响系统输出的情况，并假设采用后馈控制，通过对输入变量实施调节，使输出尽可能地接近预期的目标值。在图 1.6 中用图示加以说明。假设 N_t 表示系统中各种不确定干扰对输出的影响。在不加控制时，这些干扰会使输出从预期目标值或设定点 T 偏离。因此，尽管对过程已进行了调节，误差

$$\begin{aligned}\varepsilon_t &= Y_t - T \\ &= v(B)X_t + N_t - T\end{aligned}$$

将会在输出及目标值 T 之间出现。这里的目标是选择一个控制方程，使得误差 ε 的均方尽可能小。控制方程把 t 时刻所取的调节量 $x_t = X_t - X_{t-1}$ 表示为当前偏差 ε_t 、过去偏差 $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$ 和过去调节量 $x_{t-1}, x_{t-2}, \cdots$ 的函数。执行控制方程所要求的控制行为的装置 (人工的、电力的、气动的或电子的) 称为控制器。

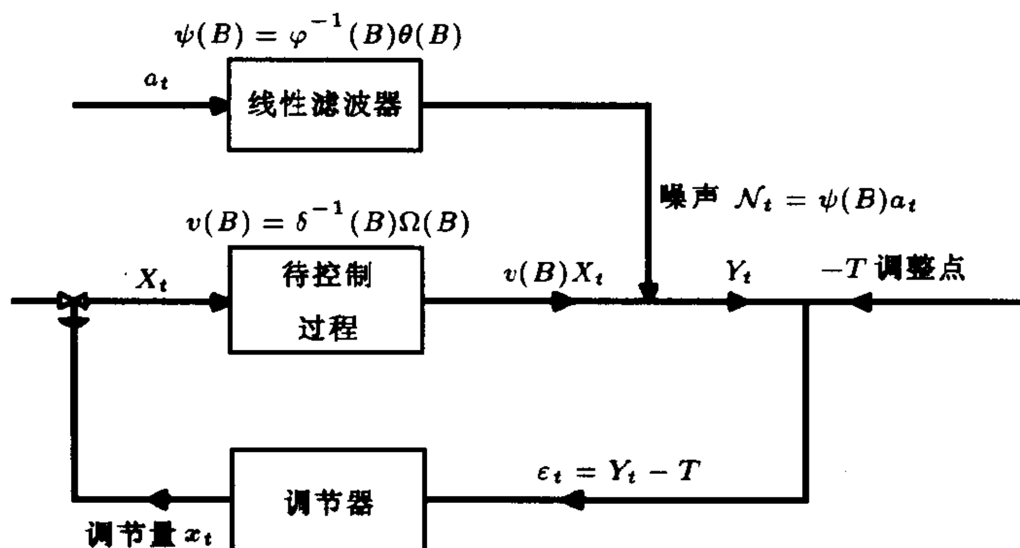


图 1.6 补偿不可观测干扰的后馈控制方案

设计控制器的一种方法就是：先预报出在不施控制时可能出现的对目标值的偏差，然后计算抵消这一偏差所需的调节量。所以预报问题和控制问题是紧密联系的。特别地，如果使用的是最小均方误差预报，则控制器将产生最小均方误差控制。为了预报不施控制时可能发生的对目标值的偏差，必须对干扰建立模型

$$N_T = \psi(B)a_t = \varphi^{-1}(B)\theta(B)a_t$$

在时刻 t, 要计算出为抵消预期输出变化所需对输入变量实施的调节量 x_t, 需要建立一个动态模型，以传递函数

$$v(B) = \delta^{-1}(B)\Omega(B)$$

把输入和输出联系起来。所得到的调节量 x_t 常由过去调节量和当前及过去控制误差的线性组合构成。这样的控制方程形式为

$$x_t = \zeta_1 x_{t-1} + \zeta_2 x_{t-2} + \cdots + \chi_0 \varepsilon_t + \chi_1 \varepsilon_{t-1} + \chi_2 \varepsilon_{t-2} + \cdots \quad (1.2.13)$$

其中 $\zeta_1, \zeta_2, \cdots, \chi_0, \chi_1, \chi_2, \cdots$ 是常数。

在实际中会出现这样的情况：最小均方误差控制有时对输出量产生极大的调节量 x_t, 以致难于接受，因此，就要使用修改的控制方案，以限制调节量的变化。这些问题将在第 13 章中讨论。

1.3 建模的基本思想

1.3.1 简约性

可以看出，我们要使用的数学模型包含某些常数或参数，其值必须由数据来估计。在实际中，对于合适的表达形式，我们应使用数目尽可能少的参数，这一点是很重要的。参数使用的简约性原则 [194] 所起的重要作用，随着我们学习的深入会越来越清楚。作为初步说明，我们考虑下面的简单例子。

假设我们拟合了一个动态模型 (1.2.9), 其形式为

$$Y_t = (\omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_s B^s)X_t \quad (1.3.1)$$

而所涉及的系统可用下式恰当地表示：

$$(1-\delta B)Y_t = \omega_0 X_t \tag{1.3.2}$$

模型 (1.3.2) 仅含两个参数： δ 和 ω_0, 但当 s 足够大时，可近似地由模型 (1.3.1)

$$Y_t = (1-\delta B)^{-1}\omega_0 X_t = \omega_0(1+\delta B+\delta^2 B^2+\cdots)X_t$$

来表示，这里 $|\delta|<1$ 。由于试验误差存在，我们很可能辨别不出拟合方程中系数之间的关系。这样我们就可能不必要地拟合一个形如 (1.3.1) 的关系式，其中包含了 $s+1$ 个参数，而简单得多的形式 (1.3.2) 仅含两个参数，它才是更适合的。因此如该例中不恰当的处理，对于给定的输入 $X_t, X_{t-1}, \cdots$ 会导致输出 Y_t 的多余且很差的估计。

因此我们的目的是必须得到适当而且简约的模型。如果模型不恰当，或参数使用冗余，将会使预报和控制严重失效，在选择模型时，慎重和反复试探非常必要。反复的选择过程是必需的，这是一个改进、修正错误和试验的过程，在后文中将简要概述。

1.3.2 模型选择的反复步骤

如果一个现象的物理结构是完全清楚的，那么写出数学表达式来准确地描述它，从理论上是可能的。由此我们得到一个结构的或理论的模型。在多数情形，完全的信息或为产生结构模型所需的大量试验手段是不可能得到的，我们只能求助于经验模型。当然，严格的结构模型和单纯的经验模型代表两种极端。实际中使用的模型常常是介于二者之间。具体地说，我们可以利用不完全的理论信息去指定一族合适的数学函数，并用来作经验拟合 (如 [40]); 也就是说，模型中所需的项数和参数的数值由试验数据来估计。这就是我们在本书中所采用的方法。当我们所描述的随机和动态模型具有正确的一般特征时，便可以对理论基础进行修正，至少可以部分地这样做。

通常假设所考虑的时间序列顺序值或输入及输出数据对于分析是有效的。如果可能，至少有 50 个，最好 100 个或更多的观测数据可以使用。在不能得到 50 个或更多的历史观测时，人们便从经验和过去的信息入手，得出一个初始模型。当得到更多的数据时，这个模型可以随时被更新。

在拟合动态模型时，有时理论分析不仅会告知我们模型的恰当形式，而且会为我们提供其参数值的很好估计。随后，通过对数据的分析可以对这些值进行检验。

图 1.7 概括了本书所使用的为了进行预报和控制建模的反复处理方法。

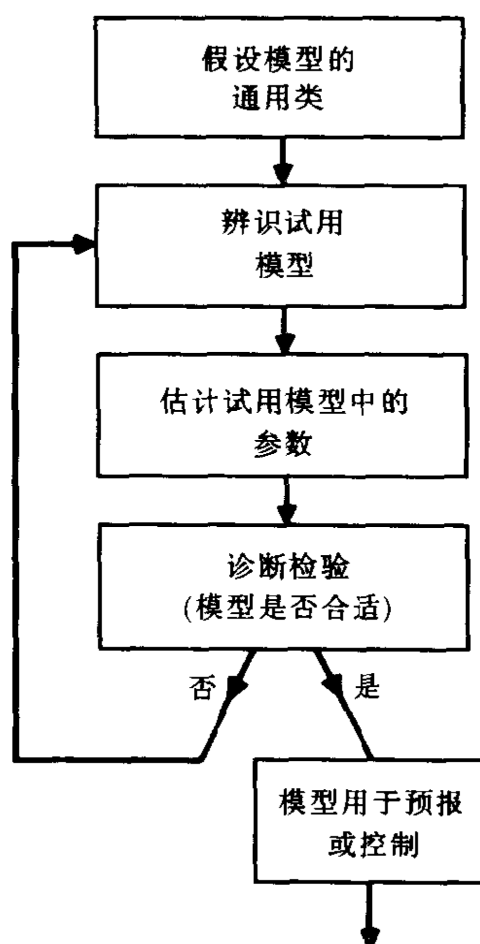


图 1.7 反复处理的建模方法步骤

1. 从理论和实际的相互作用出发，针对当前目的考虑一类有用的模型类。

2. 由于该类模型太广泛，不能方便地直接对数据进行拟合，因此提出识别这些模型子类的粗略方法，这些模型识别方法利用数据和系统信息给出模型的适当而简约的子类，以待试用。另外，识别过程还可以用来得出模型粗略的初估计。

3. 对试用模型拟合数据并估计其参数。在识别阶段得到的粗估计现在可作为参数估计中更精密迭代方法的初始值。

4. 使用诊断检验来揭示可能的拟合不足，并诊断原因。如果没有拟合不足的迹象，则该模型就可以付诸使用。如果发现不当，则重复进行识别、估计和诊断检验的反复实施，直到找到适合的表达式。

对于单变量时间序列模型讨论的识别、估计和诊断检验在第6,7,8和第9章中加以讨论，第11章讨论传递函数模型。

第一部分

随机模型及其预报

本书的第一部分包括第 2,3,4 和第 5 章，这一部分的内容对随机模型的分类作了描述，并讨论其在预报中的应用。

我们称描述观测序列概率结构的模型为随机过程。具有 N 个顺次观测值 $\boldsymbol{Z}' = (z_1, z_2, \cdots, z_N)$ 的时间序列，被看作是无限样本总体的一个实现. 我们进行统计研究的主要目的之一就是根据这些样本去推断总体的性质。例如，进行预报就是在给出过去的样本值 $\boldsymbol{Z}$ 的条件下，去推断总体中将来观测值的概率分布。为此，我们需要描述随机过程和时间序列的手段，也需要能够描述实际状况的随机过程类。在第 2 章中讨论的一类重要的随机过程是平稳过程。平稳过程被认为是统计均衡的一种特殊形式，在许多场合，特别假定过程是围绕一个固定的均值变动。描述平稳过程性状的常用手段是相关函数和谱。

在时间序列建模中一类有价值的特殊平稳随机过程是自回归过程、滑动平均过程，以及自回归滑动平均混合过程。这些过程的性质，特别是它们的自相关结构将在第 3 章中讨论。

由于许多实际中产生的时间序列 (譬如，股票价格和销售指数) 都具有非平稳特性，第 3 章中引入的平稳模型在第 4 章作了进一步推广，给出了一类有用的非平稳过程，我们称之为自回归求和滑动平均模型 (ARIMA)。所有这些模型在时间序列预报中的应用将在第 5 章中讨论，并举例加以说明。

2

平稳过程的自相关函数和谱

在时间序列模型的发展过程中，一个重要特征就是对统计均衡关系做某种形式的假设。其中一种非常特殊的假设就是平稳性的假设 (我们稍后可以看到，这是一种过份苛刻的假设) 。通常一个平稳时间序列能够有效地用其均值、方差和自相关函数 (或者等价地用其均值、方差和谱密度函数) 加以描述。在这一章中，我们要研究这些函数的性质，尤其是自相关函数的性质，这些内容在以后各章将反复用到。

2.1 平稳模型自相关的性质

2.1.1 时间序列和随机过程

时间序列　一个时间序列是依时间顺序生成的观测值的集合。若该集合是连续的，则此时间序列称为连续的。若该集合是离散的，则称时间序列是离散的。因此从一个离散时间序列在时刻 $\tau_1, \tau_2, \cdots, \tau_t, \cdots, \tau_N$ 得到的观测值可以记作 $z(\tau_1), z(\tau_2), \cdots, z(\tau_t), \cdots, z(\tau_N)$ 。在本书中我们只讨论离散时间序列，且观测值按固定的时间间隔采样。为了进行有效的分析，我们得到 N 个顺次观测值构成的序列，用 $z_1, z_2, \cdots, z_t, \cdots, z_N$ 记在等时间间隔 $\tau_0 + h, \tau_0 + 2h, \cdots, \tau_0 + th, \cdots, \tau_0 + Nh$ 得到的观测值。在许多场合 τ_0 和 h 的值是无关紧要的，但是假若观测时刻需要确切给出，这两个值就要确定了。如果我们采用 τ_0 作为原点， h 为时间单位，通常可以把 z_t 看作是时刻 t 的观测值。

离散时间序列可由以下两种方式产生：

1. 对连续时间序列进行采样：例如图 1.2 中所示，是对一个煤气炉的连续输入和输出以 9 秒为间隔的采样。

2. 每隔一段时间采集某变量的累积量：譬如，雨量通常就是每隔一段时间 (每日或每月) 采集一次，批量生产的产量也是每隔一段时间累计一次。例如，图 2.1 给出的时间序列是由一个化工过程在 70 个顺次时段的产量所组成的。

确定性时间序列和统计时间序列　若一个时间序列的未来值被某一数学函数严格确定，例如：

$$z_t = \cos(2\pi f t)$$

则称该时间序列为确定性的。若一个时间序列的未来值只能用概率分布加以描述则称之为非确定性的，或简称为统计时间序列。图 2.1 的一组数据给出了统计时间序列的例子。尽管该序列的图形看起来只是上下波动起伏，但仍不可能精确地预报出下一时段

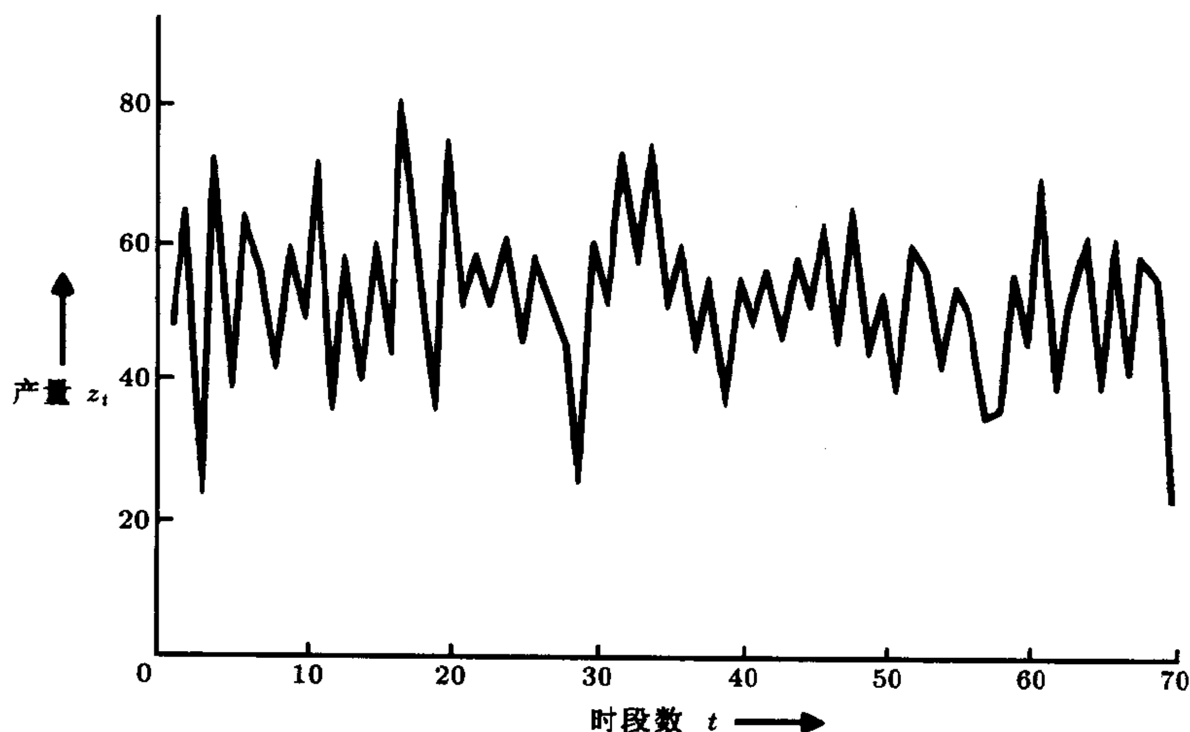


图 2.1　一个化工过程在 70 个顺次时段的产量

的产量。在本书中我们所关心的正是这种统计时间序列。

随机过程　随时间顺次发展且遵从概率法则的统计现象称之为随机过程。今后我们常常省略"随机"二字，简称为过程。我们要考查的时间序列可以看作是所研究系统的一个特殊实现，这一实现是由系统依照基本的概率机制而产生的。换言之，在考查一个时间序列时，我们将其视为某个随机过程的一个实现。例如，为了考查图 2.1 中的这批数据，不难设想，由同一化工系统在同样的 $N=70$ 个时段上可能产生许多组观测数据 (即基本随机过程的其它实现)。图 2.2 中粗线表示从 $t=21$ 到 $t=30$ 各时段的产量，图中还有其它的时间序列，它们都是从基本随机过程所定义的时间序列总体中可能得到的。由此我们可以认为在任一给定时刻 t(例如 $t=25$) 的观测值 z_t 是概率密度函数为 $p(z_t)$ 的随机变量 z_t 的一个实现。类似地，在任意两个时刻 (例如 $t_1=25, t_2=27$) 的观测值可以看作是联合概率密度函数为 $p(z_{t_1}, z_{t_2})$ 的两个随机变量 z_{t_1}, z_{t_2} 的一个实现。作为例子，在图 2.3 中给出了前文所述两变量联合分布的概率密度等值线，及其在时刻 t_1 的边际分布。一般地，由等时间间隔的观测值构成的时间序列可以用概率分布为 $p(z_1, \cdots, z_N)$ 的 $N-$ 维随机变量 $z_1, \cdots, z_N$ 来描述。也就是说，在任意的 m 个时刻 $t_1, \cdots, t_m$ 得到观测值 $z_{t_1}, z_{t_2}, \cdots, z_{t_m}$，其

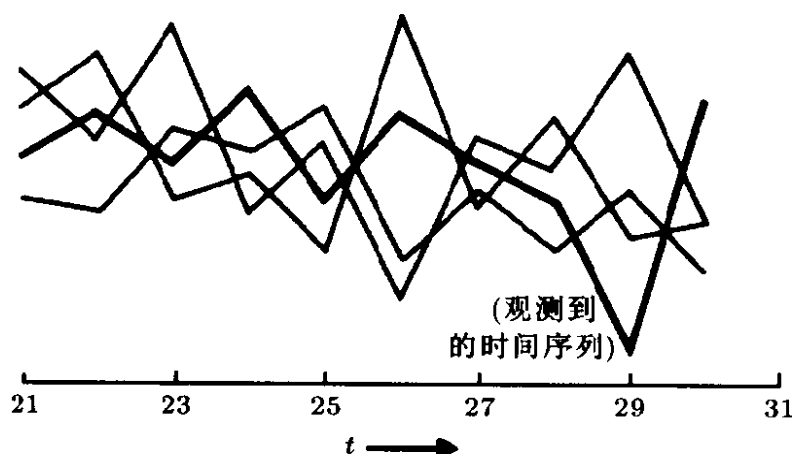


图 2.2　观测到的时间序列 (粗线)，以及表示同一随机过程其它实现的时间序列

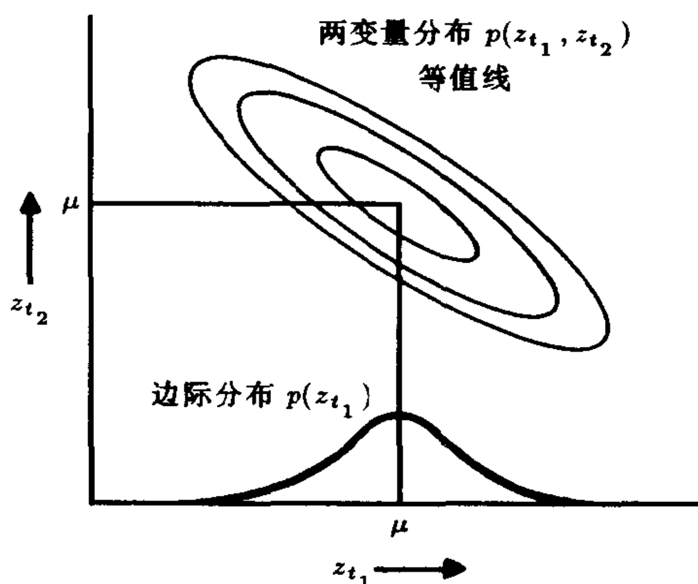


图 2.3 一个随机过程在时刻 t_1, t_2 的两变量概率分布密度等值线，以及时刻 t_1 的边际分布

联合概率分布与在时刻 $t_1+k, t_2+k, \cdots, t_m+k$ 观测值 $z_{t_1+k}, z_{t_2+k}, \cdots, z_{t_m+k}$ 的相应分布是一样的。

2.1.2 平稳随机过程

随机过程中极其特殊的一类我们称之为平稳过程，这是基于假设过程处于特定的统计均衡状态下而得到的一类过程。假若一个描述随机过程在时刻 t_1, t_2 两变量概率分布随机过程的特性不受时间原点改变的影响，我们便称其为严平稳过程。因此，对于一个具有严平稳性质的离散过程，如果把所有观测时刻都向前或向后推移任意整数 k，那么相应观测值的联合分布不受影响。

平稳过程的均值和方差 当 $m=1$ 时，平稳性假设蕴含对于任意时刻 t 概率分布 $p(z_t)$ 是相同的，可以记为 $p(z)$，因此平稳过程具有常数均值

$$\mu = E[z_t] = \int_{-\infty}^{\infty} z p(z) dz \tag{2.1.1}$$

该式定义了一个平均水准，随机过程围绕其上下波动。平稳过程

的方差也是常数

$$\sigma_z^2 = E[(z_t - \mu)^2] = \int_{-\infty}^{\infty} (z - \mu)^2 p(z) dz \tag{2.1.2}$$

该值用来度量过程在时间平均值附近的离散程度。由于在所有时刻 t 概率分布 $p(z)$ 都是相同的，因此通过对已得观测值序列 $z_1, z_2, \cdots, z_N$ 构造的直方图，可推断出分布 $p(z)$ 的形状。此外，随机过程的均值 μ 可用该时间序列的样本均值

$$\bar{z} = \frac{1}{N} \sum_{t=1}^{N} z_t \tag{2.13}$$

进行估计。随机过程的方差 σ_z^2 也可用时间序列的样本方差

$$\hat{\sigma}_z^2 = \frac{1}{N} \sum_{t=1}^{N} (z_t - \bar{z})^2 \tag{2.1.4}$$

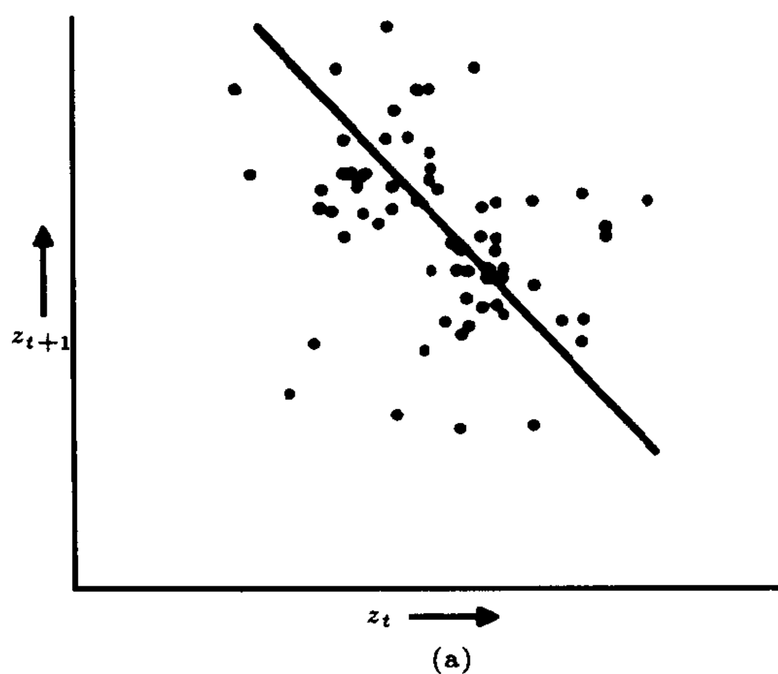
加以估计。

自协方差和自相关系数　　平稳性假定也蕴含着：对于具有常值间隔的所有时刻 t_1, t_2, 联合概率分布 $p(z_{t_1}, z_{t_2})$ 是相同的。我们按照等时间间隔 (或称滞后)k 挑出时间序列的数据对 (z_t, z_{t+k}), 并作散点图，利用散点图也可以推断联合分布的性质。对于图 2.1 中的那组数据我们作出散点图：图 2.4(a) 是滞后 $k=1$ 且 z_{t+1} 关于 z_t 的散点图；图 2.4(b) 是滞后 $k=2$ 且 z_{t+2} 关于 z_t 的散点图。我们看到，时间序列的相邻值之间是相关的， z_t 和 z_{t+1} 之间呈现负相关，而 z_t 和 z_{t+2} 之间是正相关。在平稳性假设下，若相应时间间隔为 k, 那么， z_t 和 z_{t+k} 之间的协方差对于任意 t 都是相同的，我们称之为滞后 k 的自协方差，其定义如下：

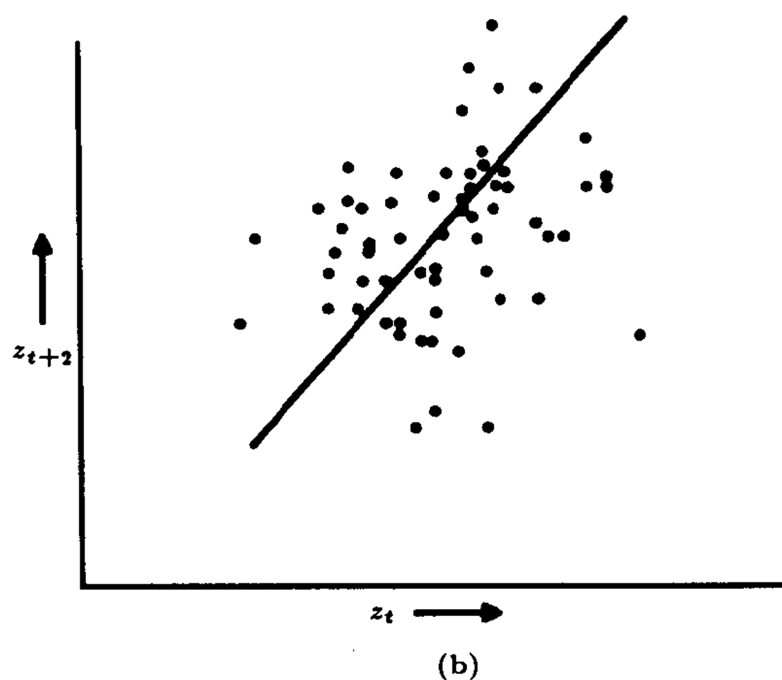
$$\gamma_k = \mathrm{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \tag{2.1.5}$$

类似地，滞后 k 的自相关函数为

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2)]}}$$
$$= \frac{E[(z_t - \mu)(z_{t+k} - \mu]}{\sigma_z^2}$$



(a)



(b)

图 2.4　图 2.1 中数据组滞后为 1 和 2 的散点图

以上结果是因为对平稳过程而言，在时刻 $t+k$ 和时刻 t 的方差都是相同的，都是 $\sigma_z^2=\gamma_0$。因此，滞后 k 的自相关 (即 z_t 和 z_{t+k} 之间的相关) 是

$$\rho_k=\frac{\gamma_k}{\gamma_0} \tag{2.1.6}$$

这蕴含 $\rho_0=1$ 。

2.1.3 正定性和自协方差阵

对一个平稳过程在 n 个顺次时刻进行取样得到的观测值 $(z_1,z_2,\cdots,z_n)$, 与其相应的协方差阵为

$$\begin{aligned}\boldsymbol{\Gamma}_n&=\begin{bmatrix}\gamma_0&\gamma_1&\gamma_2&\cdots&\gamma_{n-1}\\ \gamma_1&\gamma_0&\gamma_1&\cdots&\gamma_{n-2}\\ \gamma_2&\gamma_1&\gamma_0&\cdots&\gamma_{n-3}\\ \vdots&\vdots&\vdots&\cdots&\vdots\\ \gamma_{n-1}&\gamma_{n-2}&\gamma_{n-3}&\cdots&\gamma_0\end{bmatrix}\\ &=\sigma_z^2\begin{bmatrix}1&\rho_1&\rho_2&\cdots&\rho_{n-1}\\ \rho_1&\rho_0&\rho_1&\cdots&\rho_{n-2}\\ \rho_2&\rho_1&\rho_0&\cdots&\rho_{n-3}\\ \vdots&\vdots&\vdots&\cdots&\vdots\\ \rho_{n-1}&\rho_{n-2}&\rho_{n-3}&\cdots&1\end{bmatrix}=\sigma_z^2\boldsymbol{P}_n\end{aligned} \tag{2.1.7}$$

从上式中看到，协方差阵 $\boldsymbol{\Gamma}_n$ 是对称的，且任一对角线上的元素都相同。 $\boldsymbol{\Gamma}_n$ 称为自协方差阵，相应的阵 $\boldsymbol{P}_n$ 称为自相关阵。现在考虑随机变量 $z_t,z_{t-1},\cdots,z_{t-n+1}$ 的任一线性函数

$$L_t=l_1z_t+l_2z_{t-1}+l_nz_{t-n+1} \tag{2.1.8}$$

由于对平稳过程而言， $\text{cov}[z_i,z_j]=\gamma_{|j-i|}$, 故 L_t 的方差为

$$\text{var}[L_t]=\sum_{i=1}^{n}\sum_{j=1}^{n}l_il_j\gamma_{|j-i|}$$

若 $l_i(i=1,2,\cdots,n)$ 不全为零，则上式必然大于零。由此得知任何平稳过程的自协方差阵和自相关阵都是正定的。与此对应，如果我们把自协方差函数 $\{\gamma_k\}$ 和自相关函数 $\{\rho_k\}$ 都看作是滞后 k 的函数，对于任意正整数 n 和任意不全为零的常数 $l_1,l_2,\cdots l_n$ 应有 $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}l_il_j\gamma_{|j-i|}>0$，在这种意义下，自协方差函数和自相关函数都是正定函数。

平稳过程自相关函数需满足的条件　自相关阵 (2.1.7) 的正定性蕴含该阵的行列式及所有主子式都是大于零的。特别对于 $n=2$，有

$$\begin{vmatrix}1 & \rho_1\\ \rho_1 & 1\end{vmatrix}>0$$

因此

$$1-\rho_1^2>0$$

即

$$-1<\rho_1<1$$

类似地，对于 $n=3$ 必然有

$$\begin{vmatrix}1 & \rho_1\\ \rho_1 & 1\end{vmatrix}>0,\qquad \begin{vmatrix}1 & \rho_2\\ \rho_2 & 1\end{vmatrix}>0$$

$$\begin{vmatrix}1 & \rho_1 & \rho_2\\ \rho_1 & 1 & \rho_1\\ \rho_2 & \rho_1 & 1\end{vmatrix}>0$$

这蕴含

$$-1<\rho_1<1\,,\qquad -1<\rho_2<1$$

$$-1<\frac{\rho_2-\rho_1^2}{1-\rho_1^2}<1$$

等等。由于对所有的 n, $\boldsymbol{P}_n$ 都必然是正定的，因此平稳过程的自相关函数必须满足许多条件。在 2.2.3 中我们将会看到，所有这些条件都可以纳入谱的定义中。

线性函数的平稳性 给定 n 和系数 $l_1, l_2, \cdots, l_n$, 对于平稳过程 z_t 进行 (2.1.8) 式的线性运算得到 L_t, 由平稳性的定义推出过程 L_t 也是平稳的。特别地，一阶差分 $\nabla z_t = z_t - z_{t-1}$ 及高阶差分 $\nabla^d z_t$ 都是平稳的。对于在第 4 章中给出有关非平稳时间序列的讨论，上述结果是非常重要的。

Gauss 过程 若任意一组时间序列的观测值相应的概率分布是多维正态分布，则该过程称为正态过程或 Gauss 过程。由于多维正态分布能被它的一阶和二阶矩完全刻划，因此，只要对所有的 n 存在确定的均值 μ 和协方差阵 $\mathbf{\Gamma}_n$, 就能充分保证 Gauss 过程的平稳性。

弱平稳性 我们已经看到，对于具有严平稳性质的过程，其全部概率结构必须只依赖于时间之差。如果放宽前述限制，就是所谓的 f 阶弱平稳性 — 该过程直到 f 阶矩都是只依赖于时间之差。例如，存在确定的均值 μ 和形如 (2.1.7) 的自协方差阵就足以保证直到二阶的平稳性。因此，二阶平稳性加上正态性便导致严平稳性。

2.1.4 自协方差和自相关函数

在 2.1.2 节中我们已经看到，滞后为 k 的自相关系数 γ_k 用来度量时间间隔为 k 的两个值 z_t 和 z_{t+k} 之间的方差。如果把 γ_k 看作随滞后 k 而变化，就称 $\{\gamma_k\}$ 为随机过程的自协方差函数。类似地，把自相关系数 ρ_k 看作滞后 k 的函数，就称 $\{\rho_k\}$ 为过程的自相关函数。我们注意到，自相关函数是无量纲的，它不依赖于时间序列的计量单位。由于 $\gamma_k = \rho_k \sigma_z^2$, 若已知自相关函数 $\{\rho_k\}$ 和方差 σ_z^2 便相当于已知协方差函数 $\{\gamma_k\}$ 。

在图 2.5 中以图形方式显示自相关阵的对角线元，相关函数表现出时间序列中任意两个值之间的相关性是如何随着时间间隔而改变的。由 $\rho_k = \rho_{-k}$ 可知自相关函数必然是关于零点对称的，事实上，只需画出该函数的正半部分。图 2.6 给出了图 2.5 中自

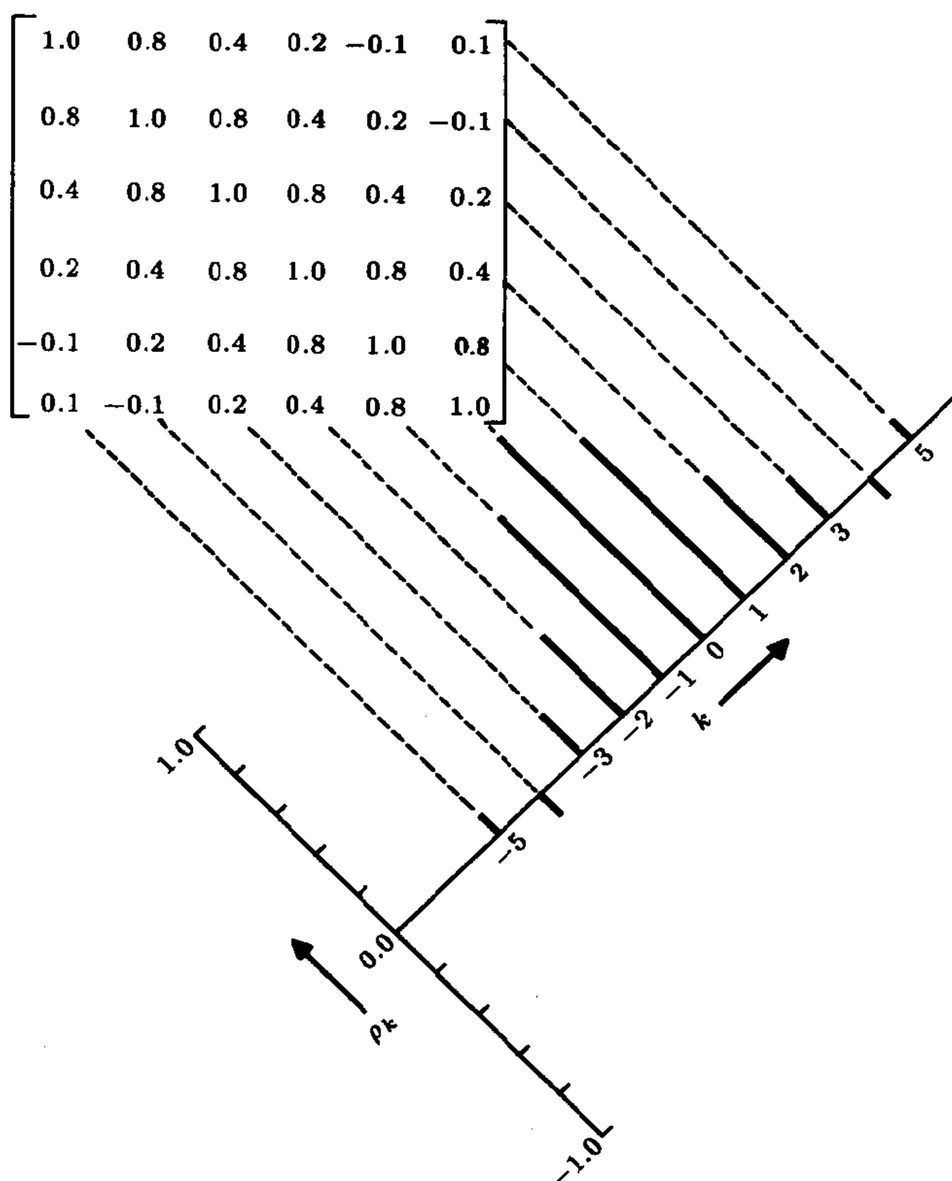


图 2.5　自相关函数及由此生成过程的自相关函数

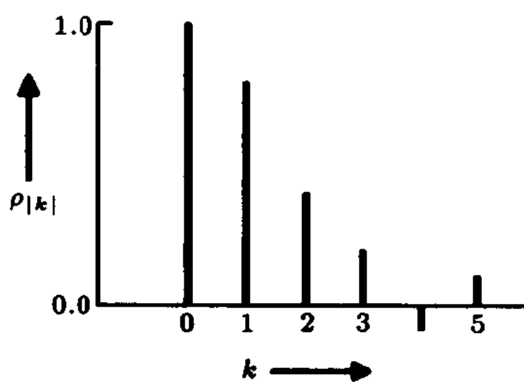


图 2.6　图 2.5 中自相关函数的正半部分

相关函数的正半部分。因此，今后当我们提到自相关函数时，通常就是指其正半部分。过去，自相关函数有时也称为相关图。

综上所述，一个正态平稳过程 z_t 能够被其均值 μ 和协方差函数 $\{\gamma_k\}$(或者等价地，均值 μ、方差 σ_z^2 和自相关函数 $\{\rho_k\}$) 完全刻划。

2.1.5 自协方差函数和自相关函数的估计

到目前为止，我们只考虑理论自相关函数，并用它来描述一个抽象的随机过程。然而在实际中，我们所具有的却是 N 个观测值 $z_1, z_2, \cdots, z_N$ 的有限时间序列，由此只能得到均值 μ 和协方差的估计值。均值 $\mu = E[Z_t]$ 由 (2.1.3) 式给出的样本均值 $\bar{z} = \sum_{t=1}^{N} z_t/N$ 来估计。容易看出 $E[\bar{z}] = \mu$，因此 $\bar{z}$ 是 μ 的无偏估计。 $\bar{z}$ 作为 μ 的一个估计量，为度量其精度我们有：

$$\text{var}[\bar{z}] = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\gamma_{t-s} = \frac{\gamma_0}{N}\left[1 + 2\sum_{k=1}^{N-1}(1 - \frac{k}{N})\rho_k\right]$$

在“大样本”情况下，上面的方差表达式可近似表示为 $\text{var}[\bar{z}] = (\gamma_0/N)(1+2\sum_{k=1}^{\infty}\rho_k)$，由此，当 $N \to \infty$ 有 $N\text{var}[\bar{z}] \to \gamma_0(1+2\sum_{k=1}^{\infty}\rho_k)$，这里假定 $\sum_{k=-\infty}^{\infty}|\rho_k| < \infty$。在 $\text{var}[\bar{z}]$ 的表达式中我们注意到，第一个因子 γ_0/N 是由容量为 N 的独立随机样本得到的 $\bar{z}$ 的方差，这一表达式是我们熟知的，但是由于 z_t 之间自相关的存在，自相关值会对 $\bar{z}$ 的精度产生严重的影响。例如，若平稳过程的自相关是 $\rho_k = \phi^{|k|}, |\phi| < 1, \bar{z}$ 方差的大样本近似式成为 $\text{var}[\bar{z}] = (\gamma_0/N)[(1+\phi)/(1-\phi)]$，第二个因子显然不是 1。

统计学家们提出了自相关函数的许多估计方法，在 [122] 中详细讨论了这些估计的性质，并认为 k 步滞后自相关 ρ_k 最令人满意的估计是

$$r_k = \frac{c_k}{c_0} \tag{2.1.9}$$

其中

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}) \qquad (k = 0, 1, 2, \cdots, K) \qquad (2.1.10)$$

是自协方差 γ_k 的估计， $\bar{z}$ 是时间序列的样本均值。(2.1.9) 式给出的 r_k 称为样本自相关函数。

为了说明 (2.1.10) 式，我们利用图 2.1 中的前 10 个数据计算 r_1 。图 2.1 的数据组在表 2.1 中给出. 表中前 10 个数据的均值 $\bar{z} = 51$, 这些数据相对于均值的偏差分别是 $-4, 13, -28, 20, -13, 13, 4, -10, 8$ 及 -3 。于是

$$\sum_{t=1}^{9} (z_t - \bar{z})(z_{t+1} - \bar{z}) = (-4)(13) + (13)(-28) + \cdots + (8)(-3)$$
$$= -1497$$

表 2.1　化工过程一组 70 个顺次产量的序列 *

1–15	16–30	31–45	46–60	61–70
47	44	50	62	68
64	80	71	44	38
23	55	56	64	50
71	37	74	43	60
38	74	50	52	39
64	51	58	38	59
55	57	45	59	40
41	50	54	55	57
59	60	36	41	54
48	45	54	53	23
71	57	48	49	
35	50	55	34	
57	45	45	35	
40	25	57	54	
58	59	50	45	

* 这一序列作为 F 序列在本书第五部分“时间序列汇集”中.

因此 $c_1 = -1497/10 = -149.7$ 。类似地，我们求出 $c_0 = 189.6$ 。故

$$r_1 = \frac{c_1}{c_0} = \frac{-149.7}{189.6} = -0.79$$

对实用而言，自相关保留两位小数足够了。上述计算只是为了进行说明。事实上，为了得到自相关函数的有效估计，至少需要 50 个观测值，待估计的自相关函数 r_k,$k = 1, 2, \cdots, K$ ，其中 K 一般不超过 $N/4$ 。

对于整个序列 (70 个观测值) 计算了 r_k 的前 15 个值，在表 2.2 中给出了这些值，并在图 2.7 中绘出图形。这些估计出的相关函数的特点是：相关性正负交替且随滞后增大而衰减。这类自相关函数对于产品数据来说实在是屡见不鲜，可能是由于“滚动”效应引起的。在这个具体例子中，若某次产量高，使得容器中存留一些生成物，因此而影响下一次的产出。

2.1.6 自相关估计的标准差

我们采用第 6 章中介绍的方法去识别某个时间序列所适合的模型，常常需要对自相关作粗略的检验，看是否超过某一滞后的 ρ_k 等效于零。为此可以利用下面的表达式，这是由 Bartlett[25] 对正态平稳过程给出的估计自相关系数方差的近似式

$$\text{var}[r_k] \simeq \frac{1}{N} \sum_{-\infty}^{+\infty} (\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2) \quad (2.1.11)$$

表 2.2 一组数据自相关函数的估计值

k	r_k	k	r_k	k	r_k
1	–0.39	6	–0.05	11	0.11
2	0.30	7	0.04	12	–0.07
3	–0.17	8	–0.04	13	0.15
4	0.07	9	–0.01	14	0.04
5	–0.10	10	–0.01	15	–0.01

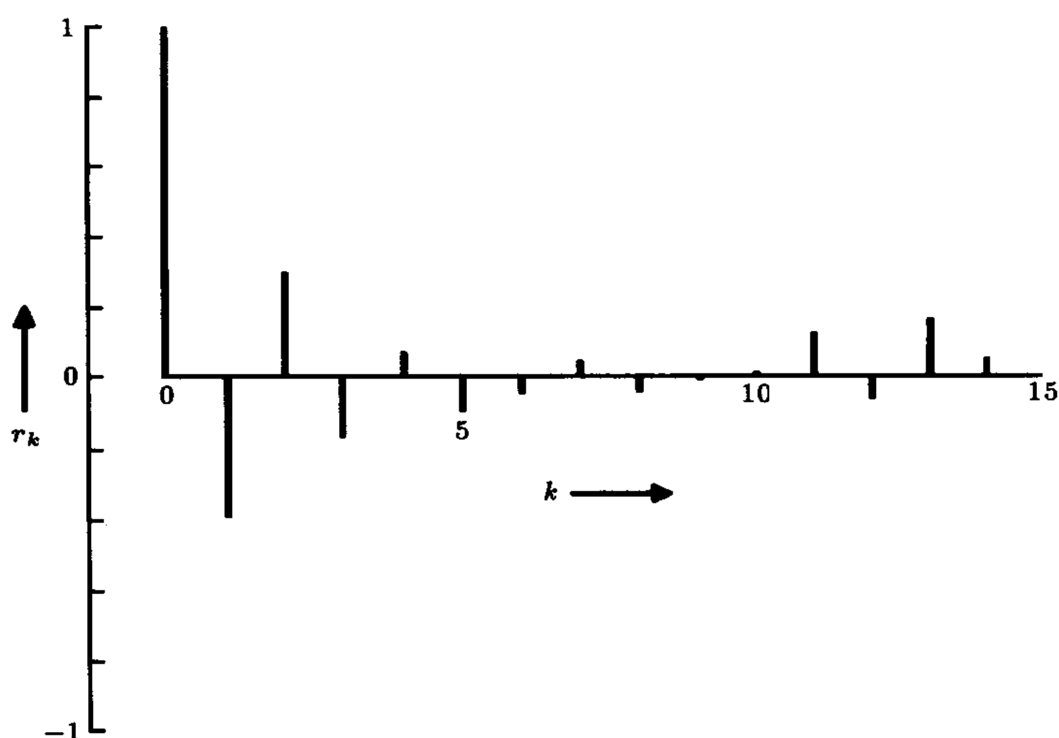


图 2.7　一组数据自相关函数的估计值

例如，若 $\rho_k = \phi^{|k|}(-1 < \phi < 1)$, 那么，自相关函数呈指数衰减，由 (2.1.11) 得

$$\mathrm{var}[r_k] \sim \frac{1}{N}\left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}\right] \tag{2.1.12}$$

特别有

$$\mathrm{var}[r_1] \simeq \frac{1}{N}(1-\phi^2)$$

对任一过程，如果当 $v > q$ 时所有的自相关 ρ_v 都是零，那么当 $k > q$ 时，在 (2.1.11) 式中除了前面几项外，其余各项都是零。因此，在计算自相关估计值的方差时，可以认为滞后 k 大于 q 的理论自相关函数" 消失 ", 由 Bartlett 近似式得到

$$\mathrm{var}[r_k] \simeq \frac{1}{N}(1 + 2\sum_{v=1}^{q}\rho_v^2) \qquad (k > q) \tag{2.1.13}$$

在实际中使用 (2.1.13) 时，是用自相关估计值 $r_k(k=1,,2,\cdots,q)$ 代替理论自相关 ρ_k，这时我们称 (2.1.13) 的平方根为大滞后标准差。假定当滞后大于 q 时理论自相关 ρ_k 全为零，那么，只要滞后足够大 $(k>q)$，大滞后标准差近似等于 r_k 的标准差。

在 [25] 中， Bartlett 还给出了不同滞后 k 和 $k+s$ 的自相关估计值 r_k 和 r_{k+s} 之间协方差的近似式，特别地，大滞后近似式化简为

$$\operatorname{cov}[r_k, r_{k+s}] = \frac{1}{N}\sum_{v=-q}^{q}\rho_v\rho_{v+s}, \qquad (k>q) \tag{2.1.14}$$

Bartlett 给出的结果 (2.1.14) 式表明，在对具体的自相关作解释时要特别小心，因为滞后邻近的自相关其协方差会很大，这些自相关之间的相互影响有时可能破坏其收敛性 (见 [122] ,[126]) 。

特别令人感兴趣的是 $q=0$ 的情形，即对于所有滞后 (除了滞后 0)ρ_k 都取值为零。因此该序列是完全随机的．于是由 (2.1.13) 得到的自相关估计 r_k 的标准差成为 $1/N^{1/2}$ 的简单形式。另外，(2.1.14) 表明，在这种情形下，不同滞后的自相关估计 r_k 和 r_{k+s} 是不相关的．由于当 N 很大时，已知 r_k 接近于正态分布，因此不同滞后自相关的全体接近于 0 均值、方差为 $1/N$ 的独立正态分布。

例 ： 一个随机过程，已知 $\rho_1=-0.4$, 且对 $k>2$ 有 $\rho_k=0$, 由该过程产生长度 N=200 的观测值时间序列，并由此得出自相关的估计值列于下表：

k	r_k	k	r_k
1	–0.38	6	0.00
2	–0.08	7	0.00
3	0.11	8	0.00
4	–0.08	9	0.07
5	0.02	10	–0.08

在序列是纯随机的假设下，应有 $q=0$, 于是由 (2.1.13) 得出对所有滞后有

$$\operatorname{var}[r_k]=\frac{1}{N}=\frac{1}{200}=0.005$$

相应的标准差是 $0.07=(0.005)^{1/2}$ 。由于 r_1 的值超出标准差的 5 倍，可以断定 ρ_1 是非零的。继而我们看到，滞后大于 1 的自相关估计值都很小，因此，下一步自然会问："ρ_1 非零，$\rho_k=0(k>2)$" 这一假设 (该假设是否适当，稍后将予以讨论) 与序列是否相合？对 $q=1$ 利用 (2.1.13) 式，并以 r_1 代替 ρ_1, 在前述假设下大滞后方差估计为

$$\operatorname{var}[r_k]\sim\frac{1}{200}[1+2(-0.38)^2]=0.0064 \qquad (k>1)$$

得到标准差是 0.08 。由于滞后大于 1 的自相关估计和标准差相比是很小的，因此，没有理由怀疑模型 $\rho_1\neq 0,\rho_k=0(k>2)$ 的合理性.

2.2 平稳模型的频谱特性

2.2.1 时间序列的周期图

时间序列可以看作是由不同频率的正弦、余弦波组成，分析时间序列的另一种方法就是建立在这一假设的基础上. 周期图便是采用这种思想的一种手段，它是由 Schuster[174] 在 1898 年引入的 (也可参见 [182]) 。周期图最初是用来检测和估计混在噪声中、频率为已知的正弦分量的振幅. 稍后我们将利用它提供一种检验序列随机性的方法 (一般是检验经模型拟合后的残差序列), 我们要考虑频率为已知的周期分量隐含于序列中的可能性.

为了说明周期图的计算方法，假设观测值的个数 $N=2q+1$ 是奇数. 我们拟合 Fourier 级数模型

$$z_t=a_0+\sum_{i=1}^{q}(a_ic_{it}+b_is_{it})+e_t \tag{2.2.1}$$

这里 $c_{it} = \cos(2\pi f_i t)$,$s_{it} = \sin(2\pi f_i t)$,$f_i = i/N$ 是基频 $1/N$ 的第 i 谐波。系数 a_0 和 (a_i, b_i) 的最小二乘估计是

$$a_0 = \bar{z} \tag{2.2.2}$$

$$a_i = \frac{2}{N}\sum_{t=1}^{N} z_t c_{it} \tag{2.2.3}$$

$$b_i = \frac{2}{N}\sum_{t=1}^{N} z_t s_{it} \tag{2.2.4}$$

其中 $i = 1, 2, \cdots, q$ 。因此周期图由 $q = (N-1)/2$ 个值组成:

$$I(f_i) = \frac{N}{2}(a_i^2 + b_i^2) \qquad (i = 1, 2, \cdots, q) \tag{2.2.5}$$

这里, $I(f_i)$ 称为在频率 f_i 处的强度。

当 N 是偶数时, 我们令 $N = 2q$, 并对 $i = 1, 2, \cdots, (q-1)$ 运用 (2.2.2),(2.2.3),(2.2.4) 及 (2.2.5), 但是

$$a_q = \frac{1}{N}\sum_{t=1}^{n}(-1)^t z_t$$
$$b_q = 0$$

及

$$I(f_q) = I(0.5) = N a_q^2$$

我们注意到, 最高的频率是每个时间间隔 0.5 周, 这是因为最小的时间间隔是 2 个时间单位。

2.2.2 方差分析

在拟合回归方程 (2.2.1) 的方差分析表中, N 为奇数时, 我们可分离出除均值外的 $(N-1)/2$ 对自由度,它们对应着参数对 $(a_1, b_1), (a_2, b_2), \cdots, (a_q, b_q)$, 从而也对应着频率 $1/N, 2/N, \cdots, q/N$ 。

周期图 $I(f_i) = (N/2)(a_i^2 + b_i^2)$ 可以简单地看成是相关参数对 (a_i, b_i)(也是相关频率) 的“平方和”。从而

$$\sum_{t=1}^{N}(z_i - \bar{z})^2 = \sum_{i=1}^{q} I(f_i) \tag{2.2.6}$$

当 N 是偶数时，有 $(N-2)/2$ 对自由度，还另有一个与系数 a_q 相应的自由度。

如果时间序列是纯随机的，不含正弦分量，即

$$z_t = \alpha_0 + e_t$$

这里 α_0 是确定的均值， $e_t(t=1,2,\cdots,N)$ 是 0 均值、方差为 σ^2 的独立正态随机变量，每一分量 $I(f_i)$ 的期望为 $2\sigma^2$, 服从 $\sigma^2\chi^2(2)$ 分布[①]，且与其它分量独立。反之，若序列包含频率 f_i, 振幅 A, 相位角 F 的正弦分量，便有

$$z_t = \alpha_0 + \alpha\cos(2\pi f_i t) + \beta\sin(2\pi f_i t) + e_t$$

这里 $A\sin F = \alpha$ 且 $A\cos F = \beta$, 此时平方和 $I(f_i)$ 将会陡增，因为 $I(f_i)$ 的预期值是 $2\sigma^2 + N(\alpha^2+\beta^2)/2 = 2\sigma^2 + NA^2/2$ 。

事实上，一个未知系统正弦分量的频率不大可能与前述频率 f_i(在这些频率上相应的强度 $I(f_i)$ 已计算出) 完全匹配。这时，在最邻近真频率处周期图将呈现增大。

表 2.3 中英格兰 1964 年的月平均温度

t	z_t	c_{1t}	t	z_t	c_{1t}
1	3.4	0.87	7	16.1	−0.87
2	4.5	0.50	8	15.5	−0.50
3	4.3	0.00	9	14.1	0.00
4	8.7	−0.50	10	8.9	0.50
5	13.3	−0.87	11	7.4	0.87
6	13.8	−1.00	12	3.6	1.00

① 可理解为，$\chi^2(m)$ 是自由度为 m 的 χ^2 随机变量，确切定义参见附录 A7.1 .

表 2.4　温度数据不同谐波正弦余弦的振幅

i	a_i	b_i
1	−5.30	−3.82
2	0.05	0.17
3	0.10	0.50
4	0.52	−0.52
5	0.09	−0.58
6	−0.30	

例：计算周期图通常需要大量观测值。然而，为了说明周期图的计算过程，我们只用 12 个数据进行计算，这是中英格兰 1964 年的月平均温度 (摄氏)，数据在表 2.3 中给出。表 2.3 还给出了计算 a_1 需要的 $c_{1t} = \cos(2\pi t/12)$，由此得到

$$a_1 = \frac{1}{6}[(3.4)(0.87) + \cdots + (3.6)(1.00)]$$
$$= -5.30$$

$a_i, b_i, i = 1, 2, \cdots, 6$ 的值在表 2.4 中给出，并得到方差分析表 2.5 。正如所料，这些温度数据的主要分量具有 12 个月的周期，也就是说，其频率为每月 1/12 周。

表 2.5　温度序列的方差分析表

i	频率 f_i	周期	周期图 I_{f_i}	自由度	均方
1	1/12	12	254.96	2	127.48
2	1/6	6	0.19	2	0.10
3	1/4	4	1.56	2	0.78
4	1/3	3	3.22	2	1.61
5	5/12	12/5	2.09	2	1.05
6	1/2	2	1.08	1	1.08
			263.10	11	23.92

2.2.3 谱和谱密度函数

为完整起见，我们还将对谱和谱密度函数作简要讨论，这一重要工具的更充分论述可参见 [122]。在本书中我们并不使用这些方法进行时间序列分析，初学者可以略去本节。

样本谱 周期图定义的 (2.2.5) 式中假定了频率 $f_i = i/N$ 是基本频率 $1/N$ 的谐波。通过引入功率谱，我们放宽了这一假定，允许频率在 0−0.5 周内连续变化。(2.2.5) 定义的周期图修改为

$$I(f) = \frac{2}{N}(a_f^2 + b_f^2) \qquad (0 < f < \frac{1}{2}) \tag{2.2.7}$$

象周期图一样，样本谱可以用来检验和估计隐含于噪声中未知频率正弦分量的振幅。特别当我们事先知道频率 f 并不具有与序列长度的谐振关系，样本谱更是实现上述目的得力工具。样本谱还为谱分析理论提供了出发点，这要用到附录 A2.1 的一个结果。该结果表明，样本谱 $I(f)$ 和协方差函数的估计值 c_k 之间是通过下面的重要关系式相联系的：

$$I(f) = 2\left[c_0 + \sum_{k=1}^{N-1} c_k \cos(2\pi f k)\right] \qquad (0 < f < \frac{1}{2}) \tag{2.2.8}$$

即样本谱是自协方差函数估计值的 Fourier 余弦变换。

谱 对于由固定频率的正弦余弦与噪声混和生成的时间序列，周期图和样本谱都是进行分析的得力工具，然而在节 2.1 中所描述的一类平稳时间序列，其特征是频率、振幅和相位都是随机变化的。这类序列的样本谱 $I(f)$ 波动剧烈，不可能给出任何有意义的解释 [122]。

然而，假设 N 个观测值的时间序列是正态平稳过程的一个实现，我们总可以由观测值计算样本谱。如前所述，这样一个过程不会有任何确定性的正弦或余弦分量，但是我们可以进行形式

上的 Fourier 分析，并对任给频率得出估计值 (a_f, b_f)，如果我们能从随机过程得到 N 个观测值的重复实现，这样就可以构造出估计值 a_f, b_f 和 $I(f)$ 的总体。由此便能计算在样本量为 N 的重复实现之下 $I(f)$ 的均值，即

$$E[I(f)] = 2\left[E[c_0] + 2\sum_{k=1}^{N-1} E[C_k]\cos(2\pi fk)\right] \tag{2.2.9}$$

对于很大的 N 可以说明 (例如 [122])，协方差系数估计 c_k 按重复实现的均值趋于理论协方差 γ_k，即

$$\lim_{N\to\infty} E[c_k] = \gamma_k$$

(2.2.9) 式对 N 趋于无穷取极限，得到功率谱 $P(f)$ 定义如下：

$$p(f) = \lim_{N\to\infty} E[I(f)] = 2\left[\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k\cos(2\pi fk)\right] \quad (0 < f < \frac{1}{2}) \tag{2.2.10}$$

我们注意到，由于

$$\begin{aligned} |p(f)| &< 2\left[|\gamma_0| + 2\sum_{k=1}^{\infty}|\gamma_k||\cos(2\pi fk)|\right] \\ &< 2\left(|\gamma_0| + 2\sum_{k=1}^{\infty}|\gamma_k|\right) \end{aligned} \tag{2.2.11}$$

因此功率谱收敛的一个充分条件是： γ_k 衰减得足够快，使得序列 (2.2.11) 收敛。由于功率谱是自协方差函数的余弦变换，所以在数学上自协方差的信息与功率谱的信息是等价的。从现在起，我们把功率谱简称为谱。

将 (2.2.10) 从 0 到 1/2 积分，得到过程 z_t 的方差为

$$\gamma_0 = \sigma_z^2 = \int_0^{1/2} p(f)\,df \tag{2.2.12}$$

(2.2.6) 式表明，通过周期图 $I(f)$ 刻划出由正弦和余弦叠加生成的序列其方差是如何在不同谐振频率上分布的，与此类似，谱 $p(f)$ 刻划了随机过程的方差如何在频率的连续范围上分布。我们可以把 $p(f)$ 解释为在频率范围 f 到 $f+df$ 上过程方差的近似度量。此外，(2.2.10) 定义了基于协方差函数 $\{\gamma_k\}$ 的谱表示，由此可以得到 $\gamma_k=\int_0^{1/2}\cos(2\pi fk)p(f)\,df$，该式和 (2.2.10) 式一起描述了过程的谱和自协方差函数之间的一一对应关系。反之，若 $\{\gamma_k\}$ 表示一个正定序列，且 (2.2.11) 中的级数收敛，由 Herglotz 定理 (见 [140]) 可知：存在唯一的函数 $p(f)$ 使得 γ_k 有谱表示 $\gamma_k=\frac{1}{2}\int_{-1/2}^{1/2}e^{i2\pi fk}p(f)\,df$ 。因此，使 (2.2.11) 收敛的平稳过程，它的谱 $p(f)$ 就可以定义成这个唯一的函数。谱表示保证了谱函数的存在，并具有 (2.1.10) 右边的形式。

谱密度函数　有时，利用自相关 ρ_k 给出谱的定义 (2.2.10) 比用自协方差函数 γ_k 更方便，由此得到的函数

$$\begin{aligned}g(f)&=\frac{p(f)}{\sigma_z^2}\\&=2\left[1+2\sum_{k=1}^{\infty}\rho_k\cos(2\pi fk)\right]\qquad(0<f<\frac{1}{2})\end{aligned}\tag{2.2.13}$$

$g(f)$ 称为谱密度函数。利用 (2.2.12) 可以看到，谱密度函数具有性质

$$\int_0^{1/2}g(f)\,df=1$$

由于 $g(f)$ 也是正的，因此它和通常的概率密度函数一样，具有相同的性质。这种相似性还可以推广到它们估计值的性质，我们下文将要讨论。

谱估计　人们希望在 (2.2.10) 中，用自协方差的估计值 c_k 代替其理论值 γ_k，从而得到谱估计。由 (2.2.8) 看到，这相当于把样

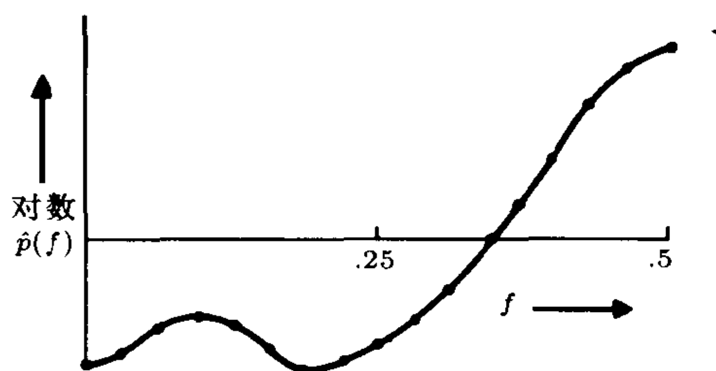


图 2.8　一组数据的功率谱估计值

本谱作为谱 $p(f)$ 的估计值。可以证明 (见 [122]) 平稳序列的样本谱围绕其理论谱剧烈波动。这一事实的直观解释是：样本谱相当于在频率域中只用了一个宽度很小的区间，这类似于在估计通常的概率分布时，只用很少的区间组去估计直方图。利用经修正或平滑的估计

$$\hat{p}(f)=2\left[c_0+2\sum_{k=0}^{N-1}\lambda_k c_k\,\cos(2\pi fk)\right] \tag{2.2.14}$$

这里 λ_k 是适当选取的权数，称为滞后窗。通过引入 (2.2.14) 式有可能增加估计值的带宽，并得到谱的较平滑的估计。(2.2.14) 中的权数 λ_k 是按一定模式选择，当滞后 $k>M$ 时应衰减到零，M 是已知的截断点，且相对于序列长度 N,M 要适当地小。用另一种计算形式，即在给定频率 f_i 附近的小邻域内，对很多的周期图作加权平均，这样就可以得到比样本谱 $I(f)$ 更平滑的谱估计。特别，$p(f)$ 的平滑周期图的估计量可取为如下形式：

$$\hat{\rho}(f_i)=\sum_{k=-m}^{m}W(f_j)I(f_i+\frac{j}{N})$$

这里 $\sum\limits_{j=-m}^{m}W(f_j)=1$ 是对称权函数，称为谱窗，m 的值应取得远比 $N/2$ 小。图 2.8 是化工产量数据的谱估计。我们看到，序列

的方差绝大部分集中在高频部分，这是由于原序列快速振荡引起的，从图 2.1 中我们可以看出这一点。

2.2.4 自协方差和谱密度函数的一些简单例子

为了说明问题，对于两个简单的随机过程我们给出如下的等价表示：

1. 理论模型
2. 理论自相关函数
3. 理论谱

考虑两个模型：

$$z_t = 10 + a_t + a_{t-1} \qquad z_t = 10 + a_t - a_{t-1}$$

这里 $a_t, a_{t-1}, \cdots$ 是均值为 0, 方差为 1, 不相关的正态随机变量序列，也即白噪声。利用定义 (2.15), 有

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)]$$

这里， $E[z_t] = E[z_{t+k}] = 0$, 上述两个随机过程的自协方差函数是

$$\gamma_k = \begin{cases} 2.0 & , \quad k = 0 \\ 1.0 & , \quad k = 1 \\ 0\,, & , \quad k > 2 \end{cases} ; \quad \gamma_k = \begin{cases} 2.0 & , \quad k = 0 \\ -1.0 & , \quad k = 1 \\ 0 & , \quad k > 2 \end{cases}$$

从而它们的理论自相关函数是

$$\rho_k = \begin{cases} 0.5 & , \quad k = 1 \\ 0.0 & , \quad k > 2 \end{cases} ; \quad \rho_k = \begin{cases} -0.5 & , \quad k = 1 \\ 0.0 & , \quad k > 2 \end{cases}$$

利用 (2.2.13) 得理论谱密度函数是

$$g(f) = 2[1 + \cos(2\pi f)] \quad ; \quad g(f) = 2[1 - \cos(2\pi f)]$$

这两个随机过程的自相关函数，谱密度函数及样本时间序列见图 2.9 。

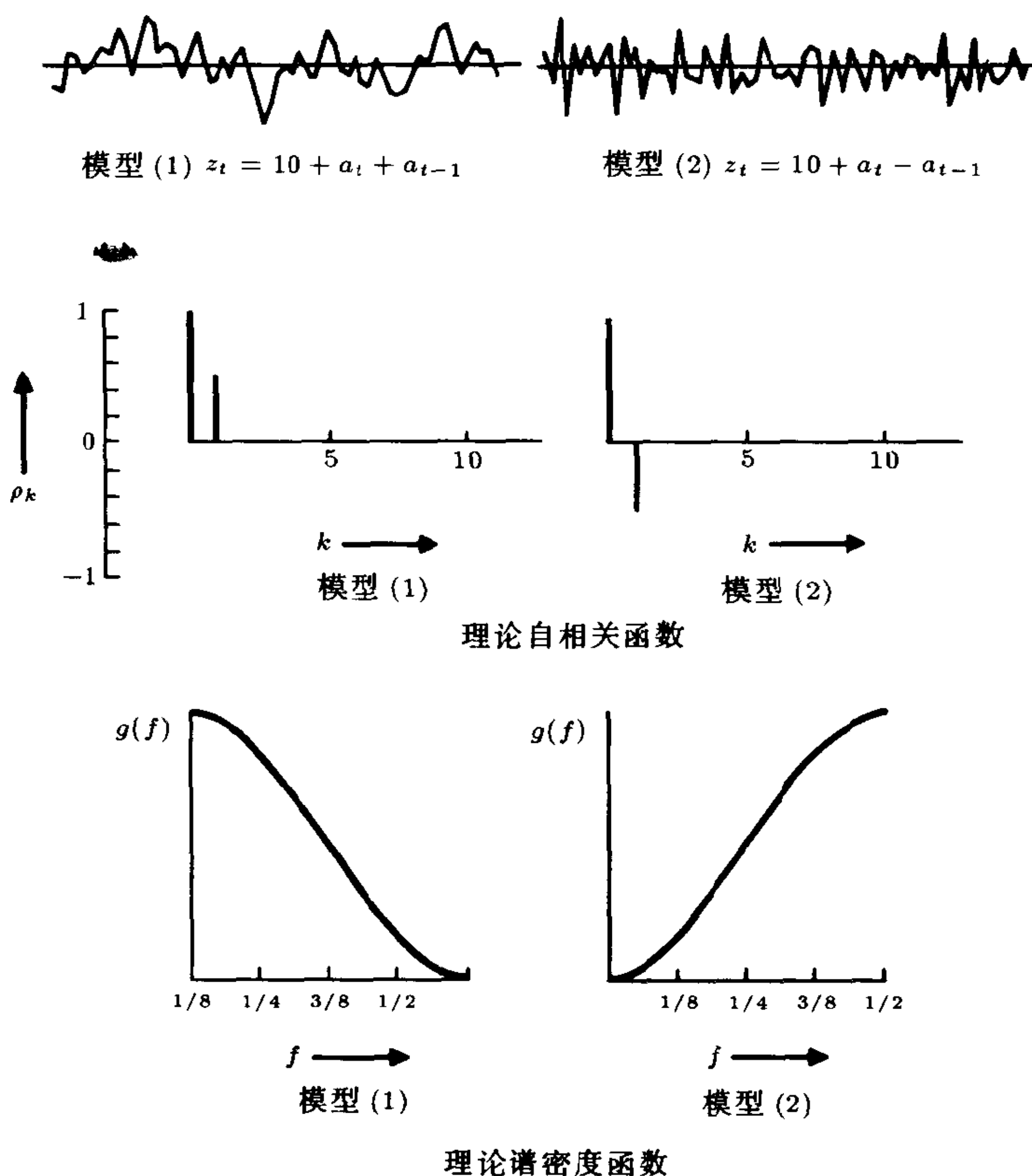


图 2.9　两个简单随机模型及对应的理论自相关函数和谱密度函数

1. 应该注意到，对于这两个平稳过程，在已知其均值和方差时，不论自相关函数或谱密度函数所提供的信息都等价于模型提供的信息 (给定正态假设)

2. 我们看到，自相关函数反映出序列某一个方面的特性。第一个序列的特征是相对平滑，这由于相邻值之间是以正相关联系。第二个序列具有正负交错的倾向，这可解释为相邻值之间以负相关联系。

3. 谱密度是从不同的角度做等价的描述,. 第一个序列的谱表现为低频占主导, 而第二序列是高频占主导.

2.2.5 自协方差和谱密度函数的优点及不足

由于自相关函数和谱可以互相转换, 在数学上是等价的, 因此有关它们优点和不足之处的讨论并不是针对数学上的等价性, 而是针对它们的表现价值. 正如我们已看到的, 每一种方法都描述了数据的不同方面, 两种方法不是对立的, 而是互补的, 都为刻划随机过程作出了贡献.

得出自相关函数和谱的估计值只是一种非结构性的研究, 这就象我们用直方图来表示一个经验分布函数. 相关估计和谱估计都是从平稳过程得到数据, 然后再去说明数据自身, 这提供了进行时间序列分析的第一步, 这也象直方图能够提供对数据作概率分布分析的第一步. 为了作进一步分析及建立参数模型, 自相关函数和谱估计是基础.

如节 2.2.4 中给出的那些参数时间序列模型并不一定就对应着简单的自相关函数或谱. 无论采用哪一种非结构方法, 我们都可能涉及多步滞后的自相关估计或是许多谱值的估计, 然而若采用参数模型, 或许只含一、两个参数的模型就能描述这些数据. 每一个自相关或谱值实际上都是一个待估计的参数, 因此, 非结构方法可能浪费许多参数, 而采用模型的方法却可能会很节省. 另一方面, 最初通常并不知道哪一类模型是适合的, 因此, 先用某种非结构方法去识别所需模型的类型, 这是十分必要的 (这就象作直方图以助于识别数据所适合的分布族) . 在建模中, 选择谱还是自相关函数作为工具取决于在实际中确有用处的那些模型的特性. 在本书后面各章将要讨论的一些模型是很有用的, 这些模型只简单地用自相关函数就能描述, 自相关函数正是我们识别模型的工具.

A2.1 样本谱和自相关函数估计之间的联系

现在我们来推导 (2.2.8) 式的结果

$$I(f) = 2\left[c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi f k)\right] \qquad (0 < f < \frac{1}{2})$$

该式把样本谱和自协方差函数的估计 c_k 联系在一起。假设在频率 f 处，上式级数中余弦和正弦分量的估计 a_f 和 b_f 是由关系式 $d_f = a_f - ib_f$ 联结的，这里 $i = \sqrt{-1}$ 。那么，

$$\begin{aligned} I(f) &= \frac{N}{2}(a_f - ib_f)(a_f + ib_f) \\ &= \frac{N}{2} d_f d_f^* \end{aligned} \qquad \text{(A.2.1.1)}$$

这里 d_f^* 是 d_f 的共轭复数。于是，利用 (2.2.3) 和 (2.2.4) 得到

$$\begin{aligned} d_j &= \frac{2}{N}\sum_{t=1}^{N} z_t[\cos(2\pi f t) - i\sin(2\pi f t)] \\ &= \frac{2}{N}\sum_{t=1}^{N} z_t e^{-i2\pi f t} \\ &= \frac{2}{N}\sum_{t=1}^{N}(z_t - \bar{z})e^{-i2\pi f t} \end{aligned} \qquad \text{(A2.1.2)}$$

将 (A2.1.2) 代入 (A2.1.1) 得到

$$I(f) = \frac{2}{N}\sum_{t=1}^{N}\sum_{t'=1}^{N}(z_t - \bar{z})(z_{t'} - \bar{z})e^{-i2\pi f(t-t')} \qquad \text{(A2.1.3)}$$

由于

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z})$$

作变换 $k = t - t'$,(A2.1.3) 成为

$$
\begin{aligned}
I(f) &= 2 \sum_{k=-N+1}^{N-1} c_k e^{-i2\pi f k} \\
&= 2\left[c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi f k)\right] \qquad (0 < f < \frac{1}{2})
\end{aligned}
$$

这正是所要的结果.

3 线性平稳模型

通常假设随机序列是由随机冲击的线性组合生成，这种随机序列就是用一般线性随机模型来描述的．在具体表达时，我们希望使用参数简约的模型．通过用数目很少的自回归和滑动平均项描述线性过程，就可以做到这一点．我们将讨论得出的自回归滑动平均 (ARMA) 模型的性质并为在建模中使用该模型作准备．

3.1 一般线性过程

3.1.1 线性过程的两种等价形式

在节 1.2.1 中我们讨论了将随机过程表示成为一个线性滤波器的输出，这里输入是白噪声 a_t，即

$$\begin{aligned}\tilde{z}_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}\end{aligned} \tag{3.1.1}$$

其中，$\tilde{z}_t = z_t - \mu$ 是过程相对某个原点的偏差，若过程是平稳的则就是相对于均值的偏差。一般线性过程允许 $\tilde{z}_t$ 表示成"白噪声"过程现在和过去值的加权和．有关线性随机模型发展的重要参考文献可参见 [25],[84],[98],[100],[126],[162],[163], [171],[180],[195],[207] 及 [212] ．白噪声过程 a_t 可以认为是驱动该系统的扰动序列，

它是由均值为零、方差为常数的不相关随机变量构成的序列。也就是

$$E[a_t]=0\ ,\qquad \mathrm{var}[a_t]=\sigma_z^2$$

由于随机变量 a_t 是不相关的，可得出它们的自协方差函数是

$$\gamma_k=E[a_t a_{t+k}]=\begin{cases}\sigma_a^2\ , & k=0\\ 0\ , & k\neq 0\end{cases}\tag{3.1.2}$$

由此得到白噪声的自相关函数具有特别简单的形式

$$\rho_k=\begin{cases}1\quad , & k=0\\ 0\quad , & k\neq 0\end{cases}\tag{3.1.3}$$

平稳过程发展中的一个基本结果是由 Wold[207] 给出的，他证明了任一零均值纯非确定的平稳过程 $\widetilde{z}_t$ 都具有形如 (3.1.1) 的线性表示，其中 $\sum_{j=0}^{\infty}\psi_i<\infty$,a_t 的方差皆为 σ_a^2, 是不相关的，但不必是独立的。对于形如 (3.1.1) 的过程 $\widetilde{z}_t$, 当 a_t 为独立随机变量时，我们仍然使用“线性过程”这一术语。

(3.1.1) 定义的 $\widetilde{z}_t$ 若表示一个有效的平稳过程，则系数 ψ_j 必须是绝对可和的，即 $\sum_{j=0}^{\infty}|\psi_j|<\infty$. 模型 (3.1.1) 表明，在适当的条件下 (见 [133,p.254]),$\widetilde{z}_t$ 也可以表示成 $\widetilde{z}$ 过去值的加权和，再加上一个附加的冲击 a_t . 即

$$\begin{aligned}\widetilde{z}_t&=\pi_1\widetilde{z}_{t-1}+\pi_2\widetilde{z}_{t-2}+\cdots+a_t\\&=\sum_{j=1}^{\infty}\pi_j\widetilde{z}_{t-j}+a_t\end{aligned}\tag{3.1.4}$$

形式 (3.1.4) 可以理解是相对于水平 μ 的当前偏差 $\widetilde{z}_t$ 关于过程过去偏差 $\widetilde{z}_{t-1},\widetilde{z}_{t-2},\cdots$ 的“回归”.

权 ψ 和权 π 之间的关系　利用前面定义的后移算子 B, 可得到权 ψ 和权 π 之间的关系。后移算子为

$$Bz_t=z_{t-1}\ ,\qquad B^j z_t=z_{t-j}$$

今后我们还需要用到前移算子 $F = B^{-1}$, 即是

$$Bz_t = z_{t+1} \ , \qquad F^j z_t = z_{t+j}$$

作为运用后移算子的例子，我们考虑模型

$$\widetilde{z}_t = a_t - \theta z_{t-1} = (1 - \theta B)a_t$$

这里 $\psi_1 = -\theta, \psi_j = 0, j > 1$ 。用 $\widetilde{z}_t, \widetilde{z}_{t-1}, \cdots$ 表示 a_t , 我们得到

$$(1 - \theta)^{-1}\widetilde{z}_t = a_t$$

因此，当 $|\theta| < 1$ 时，

$$(1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \cdots)\widetilde{z}_t = a_t$$

如同 (3.1.4) 式，将偏差 $\widetilde{z}_t$ 用过去的偏差表示，即

$$\widetilde{z}_t = -\theta\widetilde{z}_{t-1} - \theta^2\widetilde{z}_{t-2} - \theta^3\widetilde{z}_{t-3} - \cdots + a_t$$

因此，对于该模型有： $\pi_j = -\theta^j$ 。

一般， (3.1.1) 可以写成

$$\widetilde{z}_t = (1 + \sum_{j=1}^{\infty} \psi_j B^j)a_t$$

或

$$\widetilde{z}_t = \psi(B)a_t \tag{3.1.5}$$

这里

$$\psi(B) = 1 + \sum_{j=1}^{\infty} \psi_j B^j = \sum_{j=0}^{\infty} \psi_j B^j$$

令 $\psi_0 = 1$ 。正如节 1.2.1 所述， $\psi(B)$ 称为 $\widetilde{z}_t$ 关于 a_t 的线性滤波器的传递函数，我们也可以把它看作是权 ψ 的生成函数，因为如果把 B 简单地当成变量，它的 j 次幂的系数就是 ψ_j 。

类似地，(3.1.4) 可以写成

$$(1-\sum_{j=1}^{\infty}\pi_j B^j)\widetilde{z}_j = a_t$$

或

$$\pi(B)\widetilde{z}_t = a_t \tag{3.1.6}$$

因此

$$\pi(B) = 1-\sum_{j=1}^{\infty}\pi_j B^j$$

是权 π 的生成函数。(3.1.6) 式两边用 $\psi(B)$ 作用后，我们得到

$$\psi(B)\pi(B)\widetilde{z}_t = \psi(B)a_t = \widetilde{z}_t$$

这里，$\psi(B)\pi(B)=1$, 从而

$$\pi(B) = \psi^{-1}(B) \tag{3.1.7}$$

当已知权 ψ 时，可以利用关系式 (3.1.7) 推导出权 π 。反之亦然。

3.1.2 线性过程的自协方差生成函数

在第 6 章中，为了识别模型，基本的数据分析工具是自相关函数。因此，得到线性过程的自相关函数是很重要的。在附录 A3.1 中给出线性过程的自协方差函数为

$$\gamma_k = \sigma_a^2\sum_{j=0}^{\infty}\psi_j\psi_{j+k} \tag{3.1.8}$$

特别，在 (3.1.8) 中令 $k=0$, 我们求出其方差为

$$\gamma_0 = \sigma_z^2 = \sigma_a^2\sum_{j=0}^{\infty}\psi_j^2 \tag{3.1.9}$$

由此得出系数 ψ_j 绝对可和 (即 $\sum_{j=0}^{\infty}|\psi_j| < \infty$) 的平稳性条件，这蕴含 (3.1.9) 右边的级数收敛，因而保证了过程具有有限的方差。

得到线性过程自协方差的另一种方法是通过下面的自协方差生成函数

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k \tag{3.1.10}$$

应注意到，过程的方差 γ_0 是 $B^0 = 1$ 的系数，而 k 步滞后自协方差 γ_k 是 B^j 和 $B^{-j} = F^j$ 的系数。在附录 A3.1 中证明了下面表达式：

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 \psi(B)\psi(F) \tag{3.1.11}$$

例如，假设

$$\tilde{z}_t = a_t - \theta a_{t-1} = (1 - \theta B)a_t$$

因此 $\psi(B) = 1 - \theta B$. 代入 (3.1.11) 得

$$\begin{aligned}\gamma(B) &= \sigma_a^2(1 - \theta B)(1 - \theta B^{-1}) \\ &= \sigma_a^2[-\theta B^{-1} + (1 + \theta^2) - \theta B]\end{aligned}$$

与 (3.1.10) 比较，自协方差函数是

$$\begin{aligned}\gamma_0 &= (1 + \theta^2)\sigma_a^2 \\ \gamma_1 &= -\theta\sigma_a^2 \\ \gamma_k &= 0 \qquad (k > 2)\end{aligned}$$

今后，在生成函数中我们把 B 看成是一个变量，并假设它能够取复值. 特别是，经常需要考虑 $|B| < 1, |B| = 1$ 及 $|B| > 1$ 的不同情形，也就是对应于复值 B 位于单位圆内、圆上及圆外的不同情形。

3.1.3 线性过程的平稳性和可逆性条件

平稳性 (3.1.9) 式中的级数收敛就能保证线性过程具有有限的方差。我们在节 2.1.3 中已经看到，自协方差和自相关必须满足一定的条件才能保证过程的平稳性。对于线性过程 (3.1.1), 只要 $\sum_{j=0}^{\infty}|\psi_j| < \infty$ 这一个条件满足就够了。这个条件也可叙述为：作为权 ψ 生成函数的级数 $\psi(B)$ 对于 $|B| \leqslant 1$(即在单位圆上或圆内) 必须收敛。在附录 A3.1 中讨论了这一结果。

线性平稳过程的谱 在附录 A3.1 中证明：如果把 $B = e^{-i2\pi f}$(这里 $i = \sqrt{-1}$) 代入自协方差生成函数，就可以得到功率谱的半侧。因此，线性过程的功率谱为

$$
\begin{aligned}
p(f) &= 2\sigma_a^2\psi(e^{-i2\pi f})\psi(e^{i2\pi f}) \\
&= 2\sigma_a^2|\psi(e^{-i2\pi f})|^2 \qquad (0 \leqslant f \leqslant \frac{1}{2})
\end{aligned}
\tag{3.1.12}
$$

事实上， (3.1.12) 是一个熟知的表达式 [122], 它通过乘以系统的平方增益 $G^2(f) = |\psi(e^{-i2\pi f})|^2$, 把作为输入的白噪声谱 (取常值) 与线性系统输出的谱 $p(f)$ 联系起来。

可逆性 在前文中我们看到，若过程是平稳的，则线性过程的权 ψ 必定满足 $\psi(B)$ 在单位圆上或圆内收敛的条件。我们现在考虑适用于权 ψ 的另一约束条件，以保证所谓的可逆性。可逆性条件与平稳性条件是不相关的，这也适用于在第 4 章引入的非平稳线性模型。

为了说明可逆性的基本思想，再次考虑模型

$$\widetilde{z}_t = (1 - \theta B)a_t \tag{3.1.13}$$

把 a_t 用现在和过去的 $\widetilde{z}$ 表示， (3.1.13) 成为

$$a_t = (1-\theta B)^{-1}\widetilde{z}_t = (1+\theta B+\theta^2 B^2+\cdots+\theta^k B^k)(1-\theta^{k+1}B^{k+1})^{-1}\widetilde{z}_t$$

即

$$\widetilde{z}_t = -\theta\widetilde{z}_{t-1} - \theta^2\widetilde{z}_{t-2} - \cdots - \theta^k\widetilde{z}_{t-k} + a_t - \theta^{k+1}a_{t-k-1} \quad (3.1.14)$$

若 $|\theta| < 1$, 令 k 趋于无穷，我们得到无穷级数:

$$\widetilde{z}_t = -\theta\widetilde{z}_{t-1} - \theta^2\widetilde{z}_{t-2} - \cdots + a_t \quad (3.1.15)$$

那么, 用 (3.1.14) 式表示的模型的 π 权是 $\pi_j = -\theta^j$ 。总而言之, 有了 θ 的值，(3.1.13) 定义出一个名符其实的平稳过程。然而，如果 $|\theta| > 1$,(3.1.14) 中当前偏差 $\widetilde{z}_t$ 依赖于 $\widetilde{z}_{t-1}, \widetilde{z}_{t-2}, \cdots, \widetilde{z}_{t-k}$, 且相应权数随 k 而增大. 为避免出现这种情况, 要求 $|\theta| < 1$. 这时我们称序列是可逆的. 可以看出，上述条件等价于 $\sum\limits_{j=0}^{\infty}|\theta|^j \equiv \sum\limits_{j=0}^{\infty}|\pi_j| < \infty$, 因此，级数

$$\pi(B) = (1 - \theta B)^{-1} = \sum_{j=0}^{\infty} \theta^j B^j$$

对一切 $|B| < 1$(即在单位圆上或圆内) 收敛.

在第 6 章中我们将考虑模型的唯一性问题。我们将会看到，当 $|\theta| > 1$ 时，得出过程关于 a_t 的收敛的表达式是可能的，但是只能用 $\widetilde{z}_t, \widetilde{z}_{t+1}, \widetilde{z}_{t+2}, \cdots$ 表示 (即只能用过程的现在和将来值来表示) 。如果我们有兴趣把当前事件和过去发生的事件用适当的方式联系起来，那么，可逆性的条件是必不可少的.

一般, 若权 π_i 是绝对可和 (即 $\sum\limits_{j=0}^{\infty}|\pi_j| < \infty$), 这蕴含级数 $\pi(B)$ 在单位圆上或圆内收敛，于是，线性过程 (3.1.1) 是可逆的，且有如下表达式

$$\pi(B)\widetilde{z}_t = a_t$$

总之，若 $\sum\limits_{j=0}^{\infty}|\psi_j| < \infty$, 线性过程 (3.1.1) 是平稳的, 若 $\sum\limits_{j=0}^{\infty}|\pi_j| < \infty$, 则是可逆的，这里， $\pi(B) = \psi^{-1}(B) = 1 - \sum\limits_{j=1}^{\infty}\pi_j B^j$.

3.1.4 自回归和滑动平均过程

对于一般线性过程的表达式 (3.1.1) 和 (3.1.4), 如果它们包含无穷多项 ψ 和 π, 那么在实际应用中并不是很有用的. 现在我们考虑怎样引入参数节省且又能保持有效描述的模型.

自回归过程　考虑 (3.1.4) 的特殊情形: 只有前面 p 个权是非零的。模型可写为

$$\tilde{z}_t = \phi_1\tilde{z}_{t-1} + \phi_2\tilde{z}_{t-2} + \cdots + \phi_p\tilde{z}_{t-p} + a_t \tag{3.1.16}$$

这里, 我们用符号 $\phi_1, \phi_2, \cdots, \phi_p$ 记权参数的有限集合. (3.1.16) 定义的过程称为 p 阶自回归过程, 或简称为 AR(p) 过程. 特别, 一阶 ($p = 1$) 和二阶 ($p = 2$) 自回归模型

$$\tilde{z}_t = \phi_1\tilde{z}_{t-1} + a_t$$
$$\tilde{z}_t = \phi_1\tilde{z}_{t-1} + \phi_2\tilde{z}_{t-2} + a_t$$

在实际应用中是非常重要的。

现在我们将 (3.1.16) 写成下面的等价形式:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\tilde{z}_t = a_t$$

或

$$\phi(B)\tilde{z}_t = a_t \tag{3.1.17}$$

由 (3.1.17) 可得

$$\tilde{z}_t = \frac{1}{\phi(B)}a_t = \phi^{-1}(B)a_t$$

自回归过程可以看作是线性滤波器的输出 $\tilde{z}_t$, 相应传递函数为 $\phi^{-1}(B)$ 且输入是白噪声 a_t .

滑动平均过程　考虑在 (3.1.1) 中 ψ 权只有前 q 个为非零的情形. 这时过程写为

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{3.1.18}$$

这里，我们用符号 $\theta_1,\theta_2,\cdots,\theta_q$ 记权参数的有限集合。由 (3.1.18) 定义的过程称为 q 阶滑动平均过程[①]，简记为 MA(q)。特别，一阶 ($q=1$) 和二阶 ($q=2$) 的过程

$$\widetilde{z}_t = a_t - \theta_1 a_{t-1}$$
$$\widetilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

在实际中是特别有用的。

我们也可将 (3.1.18) 写成等价形式：

$$\widetilde{z}_t = (1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)a_t$$

或

$$\widetilde{z}_t = \theta(B)a_t \tag{3.1.19}$$

因此，滑动平均过程可以看作是线性滤波器的输出 $\widetilde{z}_t$，相应的传递函数为 $\theta^{-1}(B)$，输入是白噪声 a_t。

自回归－滑动平均混合模型　在节 3.1.1 中我们已经看到，一个有限的滑动平均模型

$$\widetilde{z}_t = a_t - \theta_1 a_{t-1} = (1-\theta_1 B)a_t \qquad (|\theta_1|<1)$$

能够写成无限的自回归过程。

$$\widetilde{z}_t = -\theta_1\widetilde{z}_{t-1} - \theta_1^2\widetilde{z}_{t-2} - \cdots + a_t$$

因此，若真实模型是 MA(1)，我们可以得到一个参数冗余的自回归模型，反之，AR(1) 使用滑动平均过程描述也将是不简约的表示。在实际中，为了节省参数，有时在模型中必须既包含自回归项也包含滑动平均项。即

$$\widetilde{z}_t = \phi_1\widetilde{z}_{t-1} + \cdots \phi_p\widetilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

[①] 如第 1 章所述，“滑动平均”一词多少有些含混，因为权数之和不是 1. 但是，我们还是引入并使用这一术语。

或

$$\phi(B)\widetilde{z}_t = \theta(B)a_t \tag{3.1.20}$$

我们称为 (p,q) 阶的自回归滑动平均混合模型，简记为 ARMA(p,q)。例如 ARMA(1,1) 过程是

$$\widetilde{z}_t - \phi_1\widetilde{z}_{t-1} = a_t - \theta_1 a_{t-1}$$

由于 (3.1.20) 可以写成

$$\begin{aligned}\widetilde{z}_t &= \phi^{-1}(B)\theta(B)a_t \\ &= \frac{\theta(B)}{\phi(B)}a_t = \frac{1-\theta_1 B - \cdots - \theta_q B^q}{1-\phi_1 B - \cdots \phi_p B^p}a_t\end{aligned}$$

自回归滑动平均混合模型也可以看成是线性滤波器的输出 $\widetilde{z}_t$，其传递函数是两个多项式 $\theta(B)$ 和 $\phi(B)$ 之比，输入为白噪声。我们注意到，由于 $\widetilde{z}_t = z_t - \mu$, $\mu = E[z_t]$ 是平稳过程的均值，因此，形如 (3.1.20) 的一般 ARMA(p,q) 过程可用原过程 z_t 表示为如下的等价形式：

$$\phi(B)z_t = \theta_0 + \theta(B)a_t \tag{3.1.21}$$

其中，模型的常数项是

$$\theta_0 = (1-\phi_1-\phi_2-\cdots-\phi_p)\mu \tag{3.1.22}$$

在以后各节中我们将讨论自回归、滑动平均及混合模型的重要性质。我们要研究它们的方差、自相关函数和谱，以及满足平稳性和可逆性时对参数赋予的条件。

3.2 自回归过程

3.2.1 自回归过程的平稳性条件

对于 AR(p) 模型

$$\widetilde{z}_t = \phi_1\widetilde{z}_{t-1} + \cdots + \phi_p\widetilde{z}_{t-p} + a_t$$

或

$$(1-\phi_1 B-\cdots-\phi_p B^p)\tilde{z}_t=\phi(B)\tilde{z}_t=a_t$$

为使过程是平稳的，必须调整参数 $\phi_1,\phi_2,\cdots,\phi_p$ 使其满足某些条件。

为说明起见，我们考查一阶自回归过程：

$$(1-\phi_1 B)\tilde{z}_t=a_t$$

我们可写成

$$\tilde{z}_t=(1-\phi_1 B)^{-1}a_t=\sum_{j=0}^{\infty}\phi_1^j a_{t-j}$$

因此

$$\psi(B)=(1-\phi_1 B)^{-1}=\sum_{j=0}^{\infty}\phi_1^j B^j \tag{3.2.1}$$

在节 3.1.3 中我们看到，为满足平稳性，当 $|B|<1$ 时，$\psi(B)$ 必须是收敛的。由 (3.2.1) 可知，对于 AR(1) 过程，ϕ_1 必须满足条件 $|\phi_1|<1$，才能保证平稳性。由于方程 $1-\phi_1 B=0$ 的根为 $B=\phi_1^{-1}$，平稳性条件的另一种等价叙述是：方程 $1-\phi_1 B=0$ 的根必须在单位圆之外。

对于一般的 AR(p) 过程 $\tilde{z}_t=\psi^{-1}(B)a_t$，我们得到

$$\phi(B)=(1-G_1 B)(1-G_2 B)\cdots(1-G_p B)$$

其中 $G_1^{-1},\cdots,G_p^{-1}$ 是 $\phi(B)=0$ 的根，用部分分式表示 $\phi^{-1}(B)$，得

$$\tilde{z}_t=\phi^{-1}(B)a_t=\sum_{i=1}^{p}\frac{K_i}{1-G_i B}a_t$$

因此，如果当 $|B|<1$ 时 $\psi(B)=\phi^{-1}(B)$ 是一个收敛级数，即权 $\psi_j=\sum_{i=1}^{p}K_i G_i^j$ 是绝对可和的，那么 AR(p) 就表示一个平稳过程。这时必须有 $|G_i|<1, i=1,2,\cdots,p$，等价地，$\phi(B)=0$ 的根必须

在单位圆外。方程 $\phi(B)=0$ 的根可以认为是多项式 $\phi(B)$ 的零点，因此，平稳性条件可表述为 $\phi(B)$ 的零点必须在单位圆外。当 $\phi(B)$ 有重零点时，也可以有同样的讨论。方程 $\phi(B)=0$ 称为过程的特征方程。另外，由关系式 $\phi(B)\psi(B)=1$, 自然可以得出过程 AR(p) 的权 ψ_j 满足下面差分方程

$$\psi_j=\phi_1\psi_{j-1}+\phi_2\psi_{j-2}+\cdots+\phi_p\psi_{j-p} \qquad (j>0)$$

这里， $\psi_0=1$, 且对 $j<0, \psi_j=0$ 。因此，权 ψ_j 很容易通过 ϕ_i 递推算出。事实上，由差分方程的基本理论可知 (将在附录 A4.1 中讨论), 对于无重根情形，若 ψ_j 满足前述差分方程则蕴含具有表达式 $\psi_j=\sum_{i=1}^{p}K_iG_i^j$ 。

由于级数

$$\pi(B)=\phi(B)=1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p$$

是有限项，因此，自回归过程的参数不需要任何约束条件就能保证其可逆性。

3.2.2 自回归过程的自相关函数和谱

自相关函数　我们不难求出关于自相关函数重要的递推关系式。对平稳自回归过程

$$\tilde{z}_t=\phi_1\tilde{z}_{t-1}+\phi_2\tilde{z}_{t-2}+\cdots+\phi_p\tilde{z}_{t-p}+a_t$$

两边同乘 $\tilde{z}_{t-k}$ 得到

$$\tilde{z}_{t-k}\tilde{z}_t=\phi_1\tilde{z}_{t-k}\tilde{z}_{t-1}+\phi_2\tilde{z}_{t-k}\tilde{z}_{t-2}+\cdots+\phi_p\tilde{z}_{t-k}\tilde{z}_{t-p}+\tilde{z}_{t-k}a_t \quad (3.2.2)$$

对 (3.2.2) 式取期望，得出下面的差分方程:

$$\gamma_k=\phi_1\gamma_{k-1}+\phi_2\gamma_{k-2}+\cdots+\phi_p\gamma_{k-p} \qquad (k>0) \qquad (3.2.3)$$

注意，当 $k>0$ 时，期望 $E[\tilde{z}_{t-k}a_t]$ 为零，这是因为 $\tilde{z}_{t-k}$ 只涉及直到 $t-k$ 时刻的冲击 a_j，故与 a_t 不相关。在 (3.2.3) 式两边除以 γ_0，可以看到自相关函数满足同样的差分方程

$$\rho_k=\phi_1\rho_{k-1}+\phi_2\rho_{k-2}+\cdots+\phi_p\rho_{k-p}\qquad k>0 \tag{3.2.4}$$

我们注意到，上式类似于过程 $\tilde{z}_t$ 自身所满足的差分方程。

假定将 (3.2.4) 记为

$$\phi(B)\rho_k=0$$

这里，$\phi(B)=1-\phi_1B-\cdots-\phi_pB^p$ 是作用于 k 的算子，而不是对 t 作用。记

$$\phi(B)=\prod_{j=1}^{p}(1-G_jB)$$

则 (3.2.4) 的通解 (例如可见附录 A4.1) 是

$$\rho_k=A_1G_1^k+A_2G_2^k+\cdots+A_pG_p^k \tag{3.2.5}$$

这里，$G_1^{-1},G_2^{-1},\cdots,G_p^{-1}$ 是下述特征方程

$$\phi(B)=1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p=0$$

的根。

为了保证平稳性，我们要求 $|G_i|<1$。在实际应用中，如果我们假定根是互异的，会出现两种情况：

1. G_i 是实根，这时在 (3.2.5) 中 $A_iG_i^k$ 随 k 增大等比例地衰减到零，我们常称之为指数衰减。

2. G_i 和 G_j 是一对共轭复根，这导致在 (3.2.5) 出现

$$D^k\sin(2\pi fk+F)$$

使得自相关函数呈衰减的正弦振荡，衰减系数 $D=|G_i|=|G_j|$，频率 f 满足 $2\pi f=\cos^{-1}[|\mathrm{Re}(G_i)|/D]$。

通常，一个平稳自回归过程的自相关函数的变化趋势是由衰减指数和衰减正弦振荡构成。

自回归参数用自相关表示：Yule-Walker 方程　在 (3.2.4) 中令 $k=1,2,\cdots,p$, 我们得到用 $\phi_1,\phi_2,\cdots,\phi_p$ 表示 $\rho_1,\rho_2,\cdots\rho_p$ 的线性方程组，即

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p
\end{aligned} \tag{3.2.6}$$

通常称为 Yule-Walker 方程 [195],[212]．用自相关估计值 r_k 代替理论自相关 ρ_k, 我们便得到参数的 Yule-Walder 估计。如果我们记

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}, \quad \boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}, \quad \boldsymbol{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-1} & \cdots & 1 \end{bmatrix}$$

(3.2.6) 中用自相关表示的参数 $\boldsymbol{\phi}$ 的解可以写成

$$\boldsymbol{\phi} = \boldsymbol{P}_p^{-1}\boldsymbol{\rho}_p \tag{3.2.7}$$

方差　当 $k=0$ 时，在 (3.2.3) 中取期望，项 $E[z_{t-k}z_t]$ 的贡献是 $E[a_t^2]=\sigma_a^2$, 这是因为在 z_t 中于 a_t 有关的部分只是当前的冲击 a_t．因此，当 $k=0$ 时，

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \cdots + \phi_p\gamma_{-p} + \sigma_a^2$$

上式两边除以 $\gamma_0=\sigma_z^2$, 并有 $\gamma_k=\gamma_{-k}$, 故方差 σ_z^2 可以写成

$$\sigma_z^2 = \frac{\sigma_a^2}{1-\rho_1\phi_1-\rho_2\phi_2-\cdots-\rho_p\phi_p} \tag{3.2.8}$$

谱　对于 AR(p) 过程，$\psi(B)=\phi^{-1}(B)$, 且

$$\phi(B)=1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p$$

因此，利用 (3.1.2), 自回归过程的谱为

$$p(f)=\frac{2\sigma_a^2}{|1-\phi_1e^{-i2\pi f}-\phi_2e^{-i4\pi f}-\cdots-\phi_pe^{-i2\pi pf}|^2} \quad (0<f<\frac{1}{2}) \tag{3.2.9}$$

现在我们将要讨论两个特别重要的自回归过程 — 一阶和二阶自回归过程。

3.2.3 一阶自回归 (Markov) 过程

一阶自回归过程为

$$\begin{aligned}\tilde{z}_t&=\phi_1\tilde{z}_{t-1}+a_t\\&=a_t+\phi_1a_{t-1}+\phi_1^2a_{t-2}+\cdots\end{aligned} \tag{3.2.10}$$

在节 3.2.1 中已经证明，为使过程为平稳，ϕ_1 必须满足条件 $-1<\phi_1<1$ 。

自相关函数　利用 (3.2.4), 自相关函数满足一阶差分方程

$$\rho_k=\phi_1\rho_{k-1} \qquad (k>0) \tag{3.2.11}$$

并有 $\rho_0=1$ 。此方程的解为

$$\rho_k=\phi_1^k \qquad (k>0) \tag{3.2.12}$$

如图 3.1 所示，当 ϕ_1 为正时，自相关函数指数衰减到零，当 ϕ_1 为负时，是正负交错地指数衰减到零。特别应注意到

$$\rho_1=\phi_1 \tag{3.2.13}$$

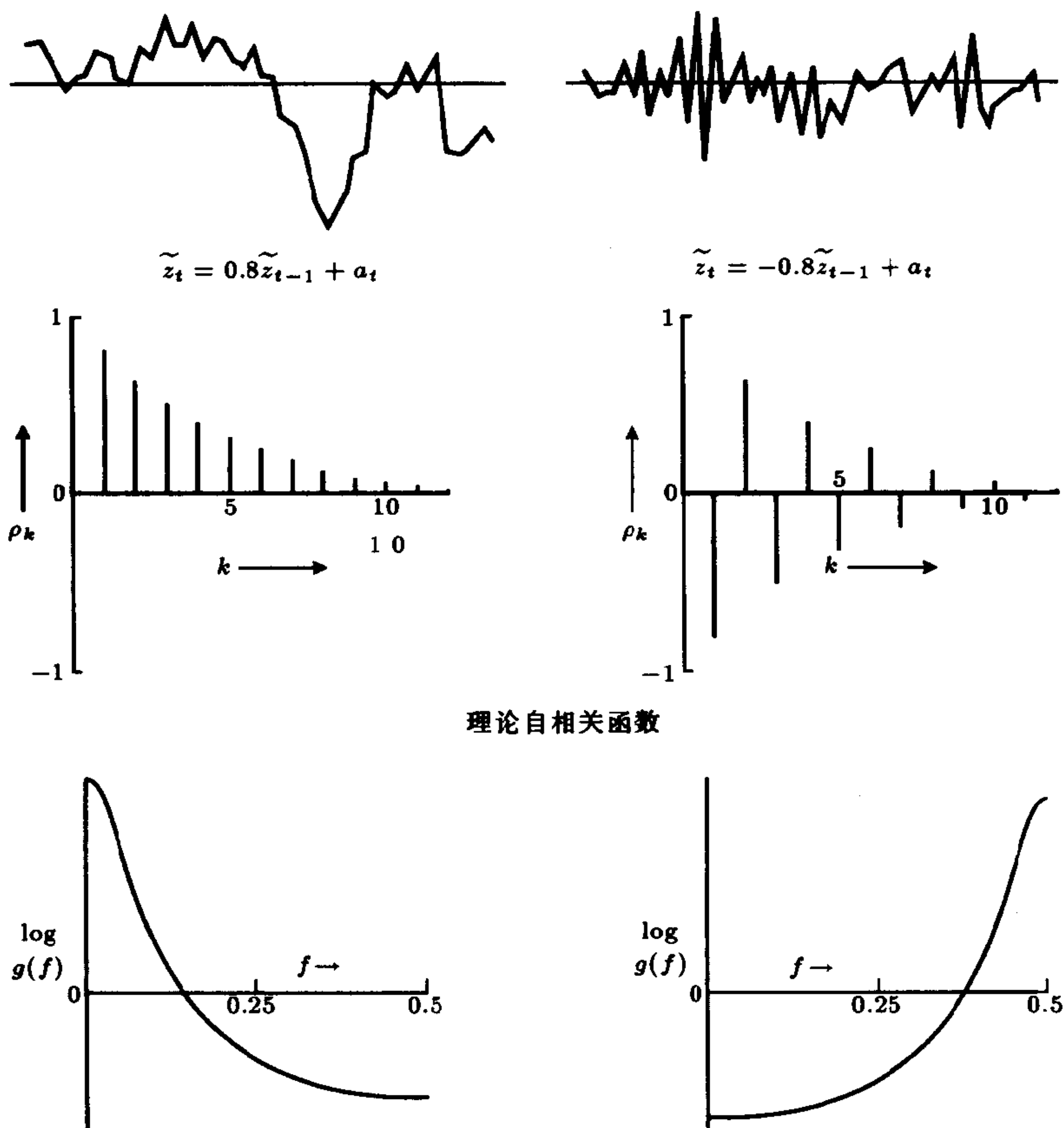


图 3.1 一阶自回归过程的实现及相应的理论自相关和谱密度函数

方差 利用 (3.2.8), 过程的方差是

$$\sigma_z^2 = \frac{\sigma_a^2}{1-\rho_1\phi_1} = \frac{\sigma_a^2}{1-\phi_1^2} \tag{3.2.14}$$

上式中用到 $\rho_1 = \phi_1$.

谱　最后，利用 (3.2.9) 得到过程的谱是

$$
\begin{aligned}
p(f) &= \frac{2\sigma_a^2}{|1-\phi_1 e^{-i2\pi f}|^2} \\
&= \frac{2\sigma_a^2}{1+\phi_1^2 - 2\phi_1\cos(2\pi f)} \quad (0 < f < \frac{1}{2})
\end{aligned}
\tag{3.2.15}
$$

图 3.1 中所示为 $\phi_1 = 0.8$ 和 $\phi_1 = -0.8$ 时过程的实现，图中并给出了相应的理论自相关和谱．当参数具有较大的正值 $\phi_1 = 0.8$ 时，序列的相邻值是接近的，且呈现明显的趋势．在自相关函数中反映为缓慢地指数衰减到零，在谱密度中则是低频占主导．另一方面，当参数具有较大的负值 $\phi_1 = -0.8$ 时，序列呈现快速振荡，在自相关函数中反映为正负交替地衰减到零，在谱密度中则是高频占主导．

3.2.4 二阶自回归过程

平稳性条件　二阶自回归过程为

$$
\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \tag{3.2.16}
$$

为保证平稳性，方程

$$
\phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0 \tag{3.2.17}
$$

的根必须在单位圆外．这蕴含参数 ϕ_1 和 ϕ_2 必须在下面三角形中：

$$
\begin{aligned}
\phi_2 + \phi_1 &< 1 \\
\phi_2 - \phi_1 &< 1 \\
-1 < \phi_2 &< 1
\end{aligned}
\tag{3.2.18}
$$

具体图形在图 3.2 中给出．

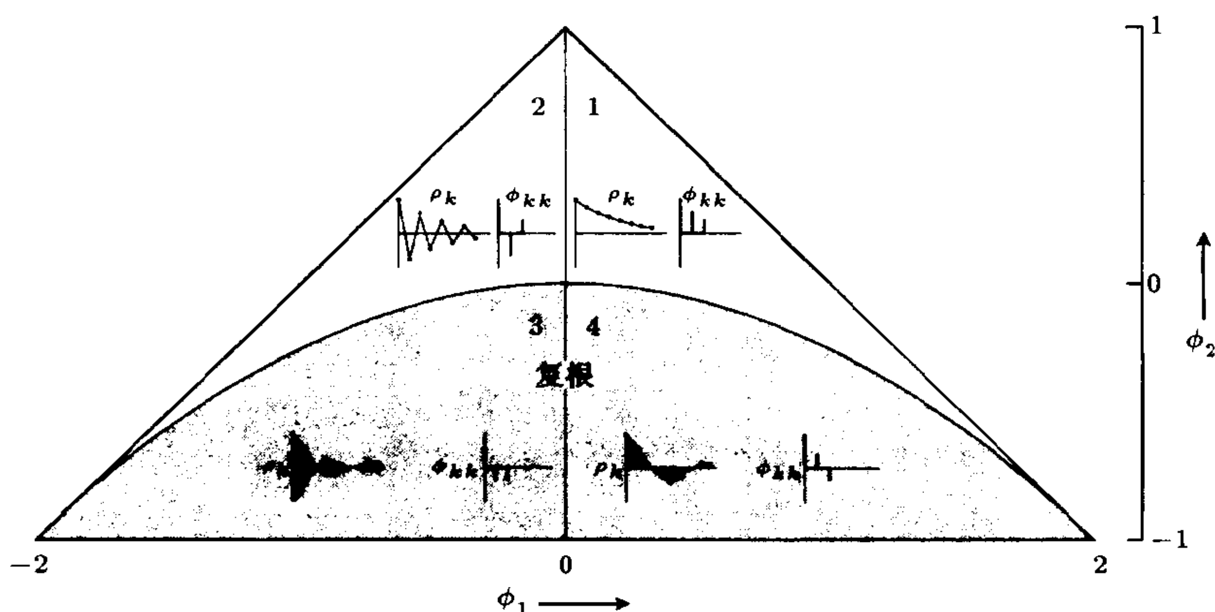


图 3.2 各种平稳 AR(2) 模型的典型自相关函数 ρ_k 和偏相关函数 ϕ_{kk}(引自 [183])

自相关函数　利用 (3.2.4), 自相关函数应满足差分方程

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \qquad (k > 0) \tag{3.2.19}$$

其中初始值 $\rho_0 = 1$, 且有 $\rho_1 = \phi_1/(1-\phi_2)$ 。由 (3.2.5) 可知, 差分方程 (3.2.19) 的通解为

$$\begin{aligned}\rho_k &= A_1G_1^k + A_2G_2^k \\ &= \frac{G_1(1-G_2^2)G_1^k - G_2(1-G_1^2)G_2^k}{(G_1-G_2)(1+G_1G_2)}\end{aligned} \tag{3.2.20}$$

这里, G_1^{-1} 和 G_2^{-1} 是特征方程 (3.2.17) 的根。若两根皆为实根, 自相关函数由两个指数衰减项生成, 这种情况出现在 $\phi_1^2+4\phi_2 > 0$ 时, 在图 3.2 中对应于抛物线上方的区域 1 和 2. 特别, 在区域 1 中, 自相关函数在衰减中保持正值, 在 (3.2.20) 式中对应正的主根。在区域 2 中, 自相关函数符号交替地衰减, 在 (3.2.20) 中对应负的主根.

当 G_1 和 G_2 为复根时 $(\phi_2^2 + 4\phi_2 < 0)$, 二阶自回归过程表现出伪周期的特征。这一特征在自相关函数中有所体现。令 $G_1 =$

$De^{i2\pi f_0}$ 和 $G_2 = De^{-i2\pi f_0}$ 并代入 (3.2.20), 我们得到

$$\rho_k = \frac{[\text{sgn}(\phi_1)]^k D^k \sin(2\pi f_0 k + F)}{\sin F} \tag{3.2.21}$$

其中 $\text{sgn}(\phi_1)$ 取值为 +1(当 ϕ_1 为正) 或 −1(当 ϕ_1 为负)．在这两种情形我们都称 (3.2.21) 为衰减正弦振荡，相应的衰减因子为 D, 频率为 f_0, 相位为 F．这些系数与二阶自回归过程参数之间的关系如下：

$$D = |G_i| = \sqrt{-\phi_2} \tag{3.2.22}$$

上式中取正根，

$$\cos(2\pi f_0) = \frac{|\text{Re}(G_i)|}{D} = \frac{|\phi_1|}{2\sqrt{-\phi_2}} \tag{3.2.23}$$

$$\tan F = \frac{1+D^2}{1-D^2}\tan(2\pi f_0) \tag{3.2.24}$$

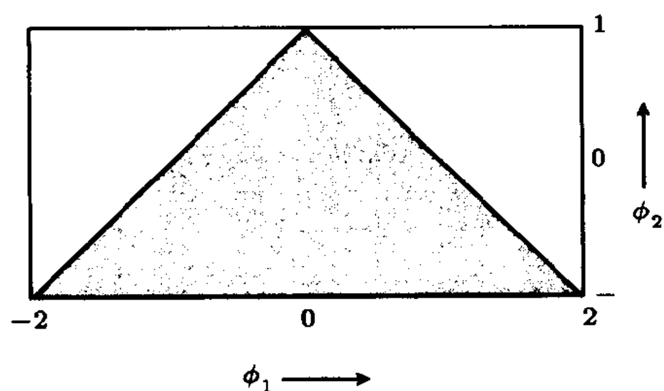
我们再次参照图 3.2, 在区域 3 和 4 中，自相关函数都是衰减正弦振荡，区域 4 中相位小于 90°，而区域 3 的相位在 90° 和 180° 之间。这意味着区域 4 中自相关函数总是从正值开始，而区域 3 中自相关函数从 0 步滞后到 1 步滞后总是会改变符号．

Yule-Walker 方程　在 (3.2.6) 中令 $p = 2$, Yule-Walker 方程成为

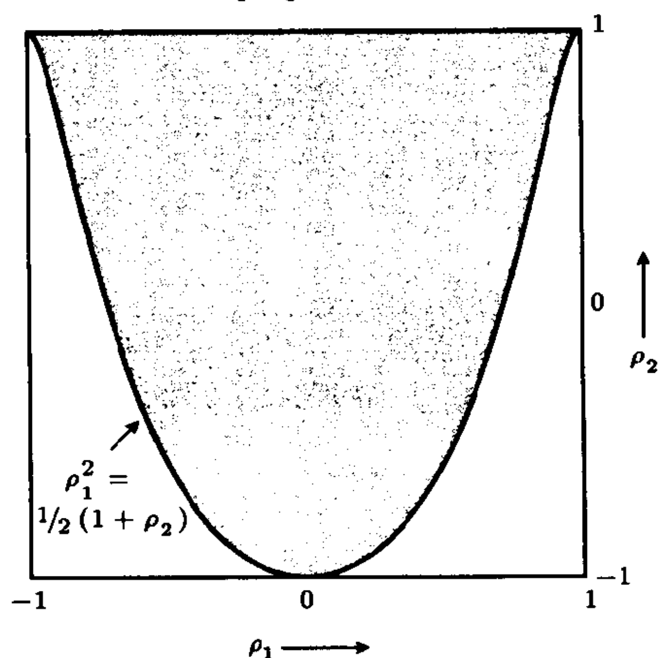
$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 \\ \rho_2 &= \phi_1\rho_1 + \phi_2 \end{aligned} \tag{3.2.25}$$

对 ϕ_1 和 ϕ_2 求解，得

$$\begin{aligned} \phi_1 &= \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\ \phi_2 &= \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{aligned} \tag{3.2.26}$$



(a) ϕ_1,ϕ_2 的允许域



(b) ρ_1,ρ_2 允许域

图 3.3 平稳 AR(2) 过程的允许域

在本书第五部分"图表汇集"的图 B 中，我们对给定的值 ρ_1 和 ρ_2 可以读出 ϕ_1 及 ϕ_2 的值。在第 6 章、第 7 章中我们将利用该图由自相关估计值 r_1 ,r_2 得出 ϕ 的估计。方程 (3.2.25) 也可对 ρ_1 和 ρ_2 求解，写成如下关于 ϕ_1,ϕ_2 的表达式

$$\rho_1=\frac{\phi_1}{1-\phi_2}\quad,\quad \rho_2=\phi_2+\frac{\phi_1^2}{1-\phi_2} \tag{3.2.27}$$

该式对前文引用的 (3.2.19) 的初值作出了说明. 自相关函数的表示形式 (3.2.20) 和 (3.2.21) 用来对实际中可能出现的各种情况进

行解释是很有用的，然而，当给定 ϕ_1,ϕ_2 时，计算 AR(2) 过程最简便的方法还是直接使用差分方程 (3.2.19)。

利用平稳性条件 (3.2.18) 及 ρ_1 和 ρ_2 的表达式 (3.2.27), 我们看到对于 AR(2) 过程，ρ_1 和 ρ_2 的允许值必须在下面区域:

$$
\begin{aligned}
-1 < \rho_1 < 1 \\
-1 < \rho_2 > 1 \\
\rho_1^2 < \frac{1}{2}(\rho_2 + 1)
\end{aligned}
$$

图 3.3.(a) 是参数 ϕ_1 和 ϕ_2 的允许域，图 3.3.(b) 是 ρ_1 和 ρ_2 相应的允许域。

方差　由 (3.2.28), 过程的方差是

$$
\begin{aligned}
\sigma_z^2 &= \frac{\sigma_a^2}{1-\rho_1\phi_1-\rho_2\phi_2} \\
&= \left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma_a^2}{\{(1+\phi_2)^2-\phi_1^2\}}
\end{aligned} \tag{3.2.28}
$$

谱　由 (3.2.9) 功率谱是

$$
\begin{aligned}
p(f) &= \frac{2\sigma_a^2}{|1-\phi_1 e^{-i2\pi f}-\phi_2 e^{-i4\pi f}|^2} \\
&= \frac{2\sigma_a^2}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos(2\pi f)-2\phi_2\cos(4\pi f)} \\
&\qquad (0 < f < \frac{1}{2})
\end{aligned} \tag{3.2.29}
$$

当特征方程具有复根时，谱也反映了序列所表现出的伪周期特性，为说明起见，图 3.4 给出的 70 个数据是由下面的二阶自回归模型产生:

$$
\tilde{z}_t = 0.75\tilde{z}_{t-1} - 0.50\tilde{z}_{t-2} + a_t
$$

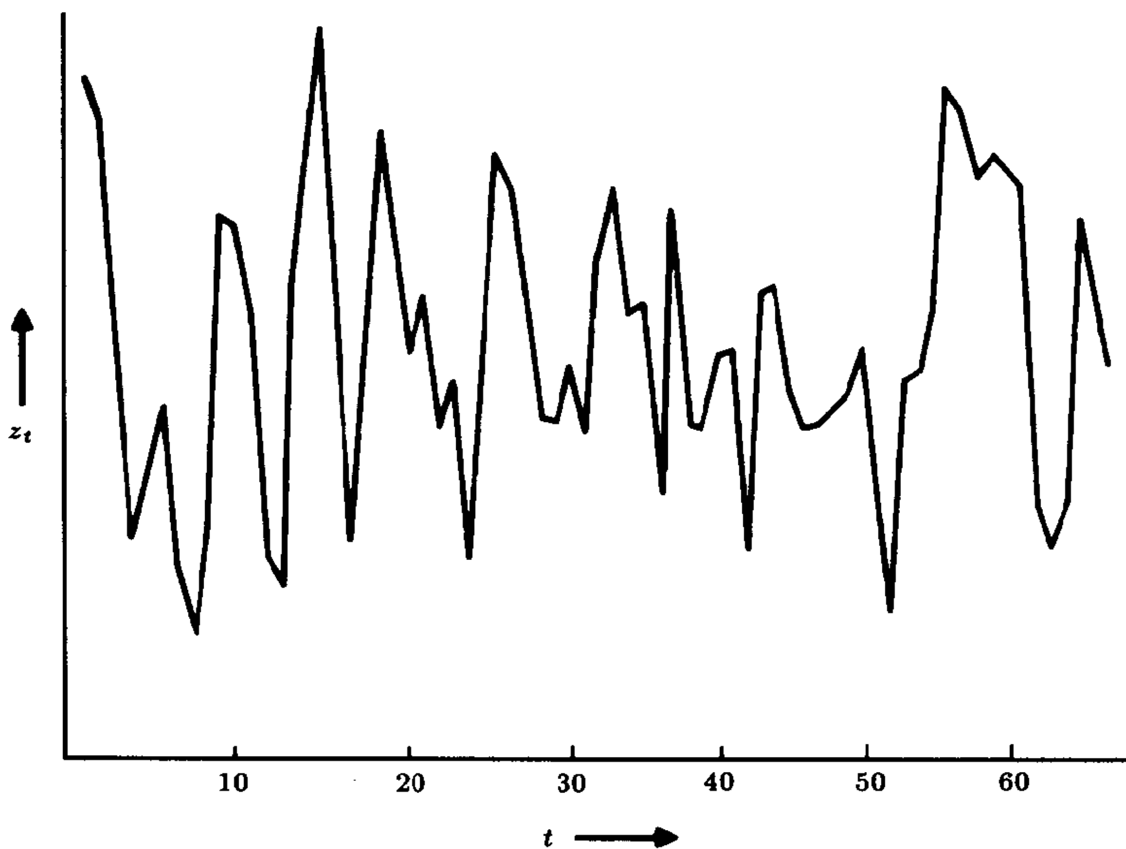


图 3.4 由二阶自回归模型 $\tilde{z}_t=0.75\tilde{z}_{t-1}-0.50\tilde{z}_{t-2}+a_t$ 产生的时间序列

此模型即是在 (3.2.16) 中令 $\phi_1 = 0.75$ 及 $\phi_2 = -0.50$ 而得到. 图 3.5 所示为由 (3.2.19) 计算的相应的理论自相关函数，初值取为 $\rho_0 = 1, \rho_1 = 0.75/[1-(-0.5)] = 0.5$ 。特征方程

$$1 - 0.75B + 0.5B^2 = 0$$

的根是复根，因此从该序列中可观察到伪周期特性是意料之中

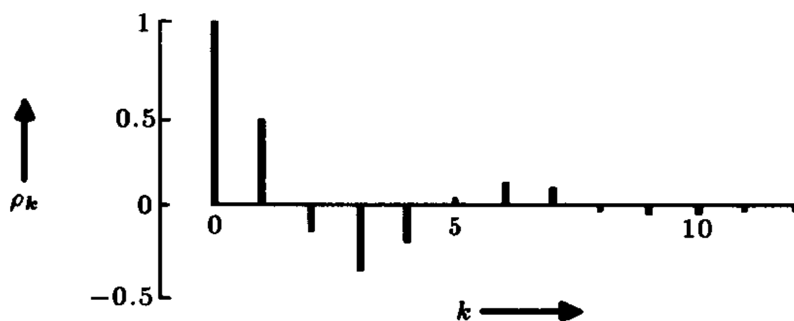


图 3.5 二阶自回归模型 $\tilde{z}_t=0.75\tilde{z}_{t-1}-0.50\tilde{z}_{t-2}+a_t$ 的理论自相关函数

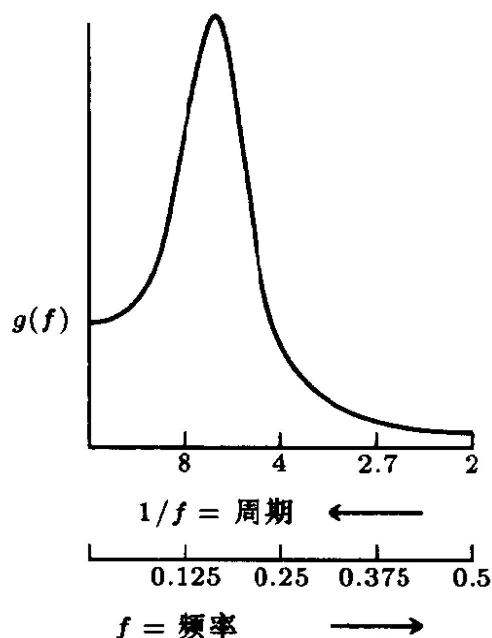


图 3.6 二阶自回归模型 $\tilde{z}_t=0.75\tilde{z}_{t-1}-0.50\tilde{z}_{t-2}+a_t$ 的理论谱密度

的。我们明显地看到，在图 3.5 的理论自相关函数中也反映了这一特性，从图上看，平均周期约为 6 。

由 (3.2.22) 和 (3.2.23) 得到衰减因子 D 和频率 f_0 为

$$D=\sqrt{0.50}=0.71 \quad , \quad f_0=\frac{\cos^{-1}(0.5303)}{2\pi}=\frac{1}{6.2}$$

因此自相关函数的基本周期为 6.2 。

最后，图 3.6 是由 (3.2.29) 得到的理论谱密度，该图表明，序列方差的大部分都集中在频率 f_0 的邻域内。

3.2.5 偏自相关函数

一开始，我们可能并不知道多少阶的自回归过程适合于观测序列。这个问题类似于在多元回归中去确定独立变量的个数。

只要一个 AR(p) 过程具有无限伸延的自相关函数，那么，就可由自相关函数的 p 个非零函数来描述其自身的特性，偏自相关函数就是基于这一事实的一种描述手段。用 ϕ_{kj} 记 k 阶回归表达式中的第 j 个系数，ϕ_{kk} 就是最后一个系数。由 (3.2.4) 知 ϕ_{kj} 满

足下面方程组

$$\rho_j = \phi_{k1}\rho_{j-1} + \cdots \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \qquad j = 1, 2, \cdots, k \tag{3.2.30}$$

由此得到 Yule-Walker 方程 (3.2.6), 可记为

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \tag{3.2.31}$$

或

$$\boldsymbol{P}_k \boldsymbol{\phi}_k = \boldsymbol{\rho}_k$$

对 $k = 1, 2, 3, \cdots$ 依次求解方程，我们得到

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \tag{3.2.33}$$

一般，对于 ϕ_{kk}, 分子和分母的行列式具有相同的元素，只是分子行列式的最后一列换成 $\boldsymbol{\rho}_k$ 。 ϕ_{kk} 看作是滞后为 k 的函数，称为偏自相关函数。

一个 p 阶自回归过程，当 k 小于或等于 p 时，偏自相关函数 ϕ_{kk} 不为零，而 k 大于 p 时， ϕ_{kk} 为零。换言之， p 阶自回归过程的偏自相关函数在滞后大于 q 时是截尾的。对于二阶自回归过程，在图 3.2 中的四个区域上都给出了相应的偏自相关函数。

对任何平稳过程，都可以由 (3.2.31) 定义偏自相关函数 ϕ_{kk}，当然也都是作为过程自相关 ρ_k 的函数，但是，对于 AR(p) 过程有：当 $k > p$ 时 $\phi_{kk} = 0$, 这是只适合于刻划 p 阶 AR 过程的明显特征。由 (3.2.31) 定义的量 ϕ_{kk} 称为过程 $\{z_t\}$ 滞后为 k 的偏自相关，这是由于 ϕ_{kk} 事实上等于 z_t 和 z_{t-k} 之间在扣除了中间变量 $z_{t-1}, z_{t-2}, \cdots z_{t-k+1}$ 的影响之后的偏相关，ϕ_{kk} 用来度量 z_t 和 z_{t-k} 之间扣除了 $z_{t-1}, z_{t-2}, \cdots z_{t-k+1}$ 影响后的相关 (或是 z_t 和 z_{t-k} 之间未被 $z_{t-1}, z_{t-2}, \cdots z_{t-k+1}$ 所解释的相关)。于是，由最小二乘原理容易得出： (3.2.31) 的解 $\phi_{k1}, \phi_{k2}, \cdots, \phi_{kk}$ 是作 z_t 关于 $z_{t-1}, \cdots, z_{t-k}$ 线性回归的回归系数，也就是 $\phi_{1k}, \phi_{2k}, \cdots, \phi_{kk}$ 是使得 $E[(z_t - b_0 - \sum_{i=1}^{k} b_i z_{t-i})^2]$ 达到极小的系数值 $b_1, \cdots, b_k$。为简便起见，假设过程 $\{z_t\}$ 具有零均值，因此，不论该过程是否为 AR 过程，在均方意义下， z_t 基于 $z_{t-1}, z_{t-2}, \cdots, z_{t-k+1}$ 的最佳线性预报 (或线性回归) 是

$$\hat{z}_t = \phi_{k-1,1} z_{t-1} + \phi_{k-1,2} z_{t-2} + \cdots + \phi_{k-1,k-1} z_{t-k+1}$$

类似地， z_t 基于 $z_{t-1}, z_{t-2}, \cdots, z_{t-k+1}$ 的最佳线性预报或线性回归是

$$\hat{z}_{t-k} = \phi_{k-1,1} z_{t-k+1} + \phi_{k-1,2} z_{t-k+2} + \cdots + \phi_{k-1,k-1} z_{t-1}$$

因此，$\{z_t\}$ 滞后为 k 的偏自相关 ϕ_{kk} 可以定义为 z_t 和 z_{t-k} 之间“扣除”之后的相关，即

$$\phi_{kk} = \text{corr}[z_t - \hat{z}_t, z_{t-k} - \hat{z}_{t-k}] \tag{3.2.34}$$

例如，我们求出 $\phi_{11} = \text{corr}[z_t, z_{t-1}]$, 而

$$\begin{aligned}\phi_{22} &= \text{corr}[z_t - \hat{z}_t, z_{t-k} - \hat{z}_{t-k}] \\ &= \frac{\gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0}{[(\gamma_0 + \rho_1^2\gamma_0 - 2\rho_1\gamma_1)^2]^{1/2}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}\end{aligned}$$

我们看到，这与前面给出 $k = 2$ 时方程 (3.2.33) 的解是一致的。可以类似地证明，由 (3.2.34) 定义的高阶偏自相关 ϕ_{kk} 是适当的

Yule-Walker 方程组 (3.2.31) 的解 (例如可见 [63,p.164]) . 由于 ϕ_{kk} 是 z_t 关于 $z_{t-1},\cdots,z_{t-k+1}$ 最小二乘回归正规方程的解，因此，ϕ_{kk} 也可看成是 z_t 关于 z_{t-k} 并扣除 $z_{t-1}\cdots,z_{t-k+1}$ 的偏回归系数。即

$$\phi_{kk}=\frac{\operatorname{cov}[z_t-\hat{z}_t,z_{t-k}-\hat{z}_{t-k}]}{\operatorname{var}[z_{t-k}-\hat{z}_{t-k}]}=\operatorname{corr}[z_t-\hat{z}_t,z_{t-k}-\hat{z}_{t-k}]$$

该式等于 $(\rho_k-\sum_{i=1}^{k-1}\phi_{k-1,i}\rho_{k-i})/(1-\sum_{i=1}^{k-1}\phi_{k-1,i}\rho_i)$, 因此

$$\sigma_{k-1}^2=\operatorname{var}[z_t-\hat{z}_t]=\operatorname{var}[z_{t-k}-\hat{z}_{t-k}]=\gamma_0(1-\sum_{i=1}^{k-1}\phi_{k-1,i}\rho_i)$$

这是由于前文用到：过程 $\{z_t\}$ 沿时间方向朝前和朝后都具有相同的相关结构。

3.2.6 偏自相关函数的估计

利用最小二乘法 (见第 7 章) 顺次拟合阶数为 1,2,3,$\cdots$ 的自回归模型，每次拟合挑出最后一个系数，得到估计值 $\hat{\phi}_{11},\hat{\phi}_{22},\hat{\phi}_{33}\cdots$, 这样就可以估计出偏自相关。另一方面，如果参数值不是十分靠近非平稳边界，那么可以采用顺次自回归模型的近似 Yule-Walker 估计。也就是，在 (3.2.30) 中用估计值 r_j 代替理论自相关得到所估计的偏自相关，即得出

$$r_j=\hat{\phi}_{k1}r_{j-1}+\hat{\phi}_{k2}r_{j-2}+\cdots+\hat{\phi}_{k(k-1)}r_{j-k+1}+\hat{\phi}_{kk}r_{j-k}\quad(j=1,2,\cdots,k)\tag{3.2.35}$$

对 $k=1,2,\cdots$ 解上述方程便得到偏自相关的估计。Durbin[87] 提出了求解方程简单的递推方法，在本书附录 A3.2 中介绍。然而，由 (3.2.35) 得出的估计值对于舍入误差是非常敏感的，因此，如果参数值接近非平稳边界，就不能使用这种递推方法。

3.2.7 偏自相关估计的标准差

Quenouille[161](也可见 [77] 和 [118]) 已证明：在 p 阶自回归过程的假设之下，阶数大于或等于 $p+1$ 的偏自相关估计值近似服从均值为零的独立正态分布。而且，若 n 是拟合时使用的观测值个数，则有

$$\text{var}[\hat{\phi}_{kk}] \simeq \frac{1}{n} \qquad (k > p+1)$$

由此，偏自相关估计 $\hat{\phi}_{kk}$ 的标准差 (S.E.) 是

$$\text{S.E.}\,[\hat{\phi}_{kk}] = \hat{\sigma}[\hat{\phi}_{kk}] \simeq \frac{1}{\sqrt{n}} \quad (k > p+1) \tag{3.2.36}$$

表 3.1 给出了对于表 2.1 数据计算的前 15 个偏自相关估计值，这是通过直接[①] 拟合顺次增加阶数的自回归模型而得到的。相应的偏自相关在图 2.1 中画出，我们可以把它们和图 2.7 中的自相关图形作比较。上述函数的特征与具有负值参数 ϕ_1 的 AR(1) 过程，或还可能与具有负主根的 AR(2) 过程 (见图 3.2 区域 2) 一致。图 3.7 中用虚线给出了 $\hat{\phi}_{22}, \hat{\phi}_{33}, \cdots$ 的 2S.E. 界限，该值是在过程为 AR(1) 的假定之下由 (3.2.26) 算出的。由于在所考虑的偏自相关中 $\hat{\phi}_{22}$ 是第二大的，因此应该注意到过程有可能是 AR(2)。

有关偏相关函数在模型识别中的应用将在第 6 章中作更充分的讨论。

表 3.1　对表 2.1 中数据的偏自相关估计值

k	$\hat{\phi}_{kk}$	k	$\hat{\phi}_{kk}$	k	$\hat{\phi}_{kk}$
1	–0.40	6	–0.15	11	0.18
2	0.19	7	0.05	12	–0.05
3	0.01	8	0.00	13	0.09
4	–0.07	9	–0.10	14	0.18
5	–0.07	10	0.05	15	0.01

[①] 与求解方程 (3.2.35) 得到的近似值，小数点后第一位一致。

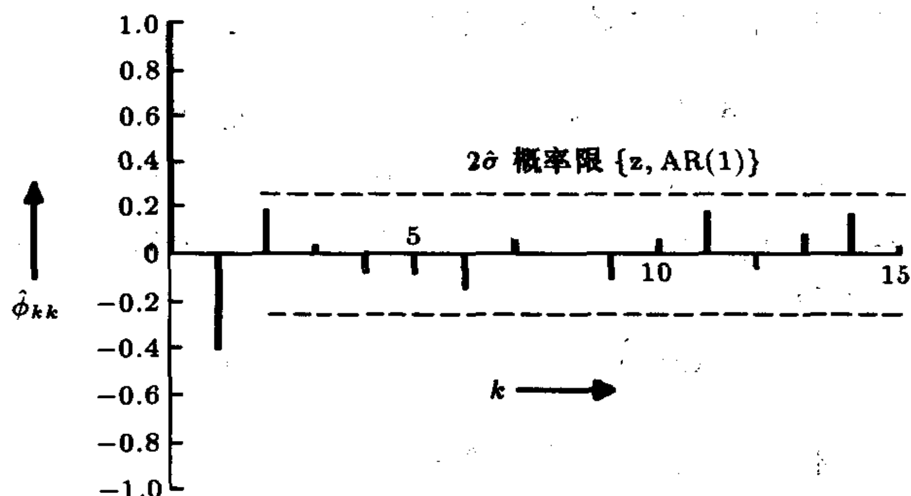


图 3.7 图 2.1 中数据的偏自相关估计值以及在 AR(1) 模型假定之下 2 倍标准差范围

3.3 滑动平均过程

3.3.1 滑动平均过程的可逆性条件

滑动平均过程 MA(q) 为

$$
\begin{aligned}
\widetilde{z}_t &= a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \\
&= (1 - \theta_1 B - \cdots - \theta_q B^q) a_t \\
&= \theta(B) a_t
\end{aligned}
\tag{3.3.1}
$$

现在我们推出为使过程 MA(q) 具有可逆性，参数 $\theta_1, \theta_2, \cdots, \theta_q$ 所必须满足的条件.

在节 3.1.3 中我们已经看到，对于一阶滑动平均过程

$$\widetilde{z}_t = (1 - \theta_1 B) a_t$$

当 $|\theta_1| < 1$ 时该过程为可逆的，即

$$\pi(B) = (1 - \theta_1 B)^{-1} = \sum_{j=0}^{\infty} \theta_1^j B^j$$

在单位圆上或圆内收敛。换言之，即是 $(1-\theta_1 B)=0$ 的根 $B=\theta_1^{-1}$ 在单位圆外。

为得到高阶 MA 过程的可逆性条件，我们将 (3.3.1) 写成

$$a_t=\theta^{-1}(B)\tilde{z}_t$$

若

$$\theta(B)=\prod_{i=1}^{q}(1-H_iB)$$

因而，用部分分式展开，我们得到

$$\pi(B)=\theta^{-1}(B)=\sum_{i=1}^{q}\left(\frac{M_i}{1-H_iB}\right)$$

当 $|H_i|<1(i=1,2,\cdots,q)$ 时，上式收敛，或等价地，权 $\pi_j=-\sum_{i=1}^{q}M_iH_i^j$ 为绝对可和。由于 $\theta(B)=0$ 的根是 H_i^{-1}，因此得知 MA(q) 过程的可逆性条件为：特征方程

$$\theta(B)=1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q=0 \qquad (3.3.2)$$

的根在单位圆外。由 $\pi(B)\theta(B)=1$ 可知权 π_j 满足差分方程

$$\pi_j=\theta_1\pi_{j-1}+\theta_2\pi_{j-2}+\cdots+\theta_q\pi_{j-q} \qquad (j>0)$$

通常取 $\pi_0=-1$ 及 $\pi_j=0$ 当 $j<0$，因此，π_j 可以很容易地由 θ_j 递推算出。

注意，由于级数

$$\psi(B)=\theta(B)=1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q$$

是有限项，为使滑动平均过程的参数满足平稳性条件，不需要有任何限制条件。

3.3.2 滑动平均过程的自相关函数和谱

自相关函数　利用 (3.3.1),MA(q) 过程的自相关函数为

$$\gamma_k = E[(a_t-\theta_1 a_{t-1}-\cdots-\theta_q a_{t-q})(a_{t-k}-\theta_1 a_{t-k-1}-\cdots-\theta_q a_{t-k-q})]$$

因此该过程的方差是

$$\gamma_0 = (1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)\sigma_a^2 \tag{3.3.3}$$

且

$$\gamma_k = \begin{cases} (-\theta_k+\theta_1\theta_{k+1}+\theta_2\theta_{k+2}+\cdots+\theta_{q-k}\theta_q)\sigma_a^2 & , k=1,2,\cdots,q \\ 0 & , k>q \end{cases}$$

由此得出相关函数是

$$\rho_k = \begin{cases} \dfrac{-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2}, & k=1,2,\cdots,q \\ 0 , & k>q \end{cases} \tag{3.3.4}$$

我们看到，对于 MA(q) 过程，当滞后超出过程的阶数 q 时自相关函数为零。换言之，滑动平均过程的自相关函数具有超出 q 步滞后的截尾性。

滑动平均参数的自相关表示　若 $\rho_1,\rho_2,\cdots,\rho_q$ 为已知，由 q 个方程 (3.3.4) 就可以解出参数 $\theta_1,\theta_2,\cdots,\theta_q$ 。然而，与自回归过程的 Yule-Walker 方程不同，方程 (3.3.4) 不是线性的，而是非线性的。 $q=1$ 是最简单的情形，稍后我们就要讨论，除此以外，有关的非线性方程组只能用附录 A6.2 中所述方法迭代求解。在 (3.3.4) 中用估计值 r_k 代替 ρ_k 并解得到的方程，我们就可以得出滑动平均参数的初估计。虽然也是通过代入 Yule-Walker 方程求解，但是与自回归估计的情况不同，这里得到的估计并不具有很好的统计有效性。尽管如此，还是为模型识别阶段提供了有用的

粗估计，有关内容在第 6 章中讨论。此外，这些估计还为第 7 章将要讨论的迭代过程提供有用的初值，以使迭代收敛到有效的极大似然估计。

谱　对于 MA(q) 过程

$$\psi(B)=\theta(B)=1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q$$

因此利用 (3.1.12); MA(q) 过程的谱是

$$p(f)=2\sigma_a^2|1-\theta_1e^{-i2\pi f}-\theta_2e^{-i4\pi f}-\cdots-\theta_qe^{-i2\pi qf}|^2 \quad (0<f<\frac{1}{2}) \tag{3.3.5}$$

下面我们将更详细地讨论一阶和二阶滑动平均过程，这些模型在实际中是非常重要的.

3.3.3 一阶滑动平均过程

我们已经见过如下形式的过程:

$$\begin{aligned}\tilde{z}_t&=a_t-\theta_1a_{t-1}\\&=(1-\theta_1B)a_t\end{aligned}$$

在节 3.1.3 中并已说明，为使过程为可逆，θ_1 必须在范围 $-1<\theta_1<1$ 内，然而，对 θ_1 的任意值，过程总是平稳的。

自相关函数　利用 (3.3.3), 过程的方差为

$$\gamma_0=(1+\theta_1^2)\sigma_a^2$$

利用 (3.3.4), 过程的自相关函数是

$$\rho_k=\begin{cases}\dfrac{-\theta_1}{1+\theta_1^2}, & k=1\\ 0\quad, & k>2\end{cases} \tag{3.3.6}$$

由上式可看到，MA(1) 过程的 ρ_1 必定满足 $|\rho_1| < \frac{1}{2}$。在 (3.3.6) 式中令 $k=1$ 得到

$$\theta_1^2 + \frac{\theta_1}{\rho_1} + 1 = 0 \tag{3.3.7}$$

该方程的关于 θ_1 的根为 $\theta_1 = (-1 \pm \sqrt{1-4\rho_1^2})/(2\rho_1)$。由于两根之积等于 1, 因此若一根为 θ_1, 则另一根为 θ_1^{-1}。而且如果 θ_1 满足可逆性条件 $|\theta_1| < 1$, 则另一根 θ_1^{-1} 必定大于 1, 即不满足可逆性条件。例如，若 $\rho_1 = -0.4$, 那么，(3.3.7) 有两个解 $\theta_1 = 0.5$ 及 $\theta_1 = 2.0$。然而，只有一个解 $\theta_1 = 0.5$ 对应于可逆过程。在第五部分“图表汇集”的表 A 中，对于可能值的全范围 $-0.5 < \rho_1 < 0.5$, 提供了对应 θ_1 解的读数。

谱 利用 (3.3.5), 谱是

$$\begin{aligned} p(f) &= 2\sigma_a^2|1-\theta_1 e^{-i2\pi f}|^2 \\ &= 2\sigma_a^2[1+\theta_1^2 - 2\theta_1\cos(2\pi f)] \qquad (0 < f < \frac{1}{2}) \end{aligned} \tag{3.3.8}$$

一般，若 θ_1 为负，ρ_1 就是正的，这时低频谱占主导。反之，当 θ_1 为正，ρ_1 便为负，高频谱占主导。

偏自相关函数 利用 (3.2.31) 及 $\rho_1 = -\theta_1/(1+\theta_1^2)$, 并有 $\rho_k = 0, k > 1$, 通过一些代数运算得

$$\phi_{kk} = \frac{-\theta_1^k(1-\theta_1^2)}{(1-\theta_1^{2(k+1)})}$$

因此 $|\phi_{kk}| < \theta_1^k$, 且偏自相关被衰减指数所控制。若 ρ_1 为正，那么 θ_1 就是负的，这时偏自相关函数的符号正负交替变换。然而，若 ρ_1 为负，则 θ_1 为正，偏自相关函数就是负的。由 (3.1.15) 可以看出，MA(1) 过程的权 π_j 为 $\pi_j = -\theta_1^j$, 由于这些权数是过程无穷自回归形式的系数，因此 MA(1) 的偏自相关函数就象权 π_j 一样具有指数衰减的特性，这是完全能理解的。

现在我们来看一下 AR(1) 和 MA(1) 过程的对偶性。MA(1) 过程的自相关函数滞后大于 1 为截尾，AR(1) 过程的自相关函数却是指数衰减式的拖尾。与此相反，MA(1) 过程的偏自相关函数为拖尾且被衰减指数控制，而过程 AR(1) 的偏自相关函数滞后大于 1 时截尾。可以证明，在一般情形下，AR 和 MA 过程的自相关和偏自相关函数也会出现这类相似的对偶关系。

3.3.4 二阶滑动平均过程

可逆性条件　二阶滑动平均过程由下式定义：

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

无论 θ_1 和 θ_2 取何值，过程都是平稳的。然而，为使过程为可逆，仅当特征方程

$$1 - \theta_1 B - \theta_2 B^2 = 0 \tag{3.3.9}$$

的根在单位圆外，也就是满足

$$\begin{aligned} \theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 < \theta_2 &< 1 \end{aligned} \tag{3.3.10}$$

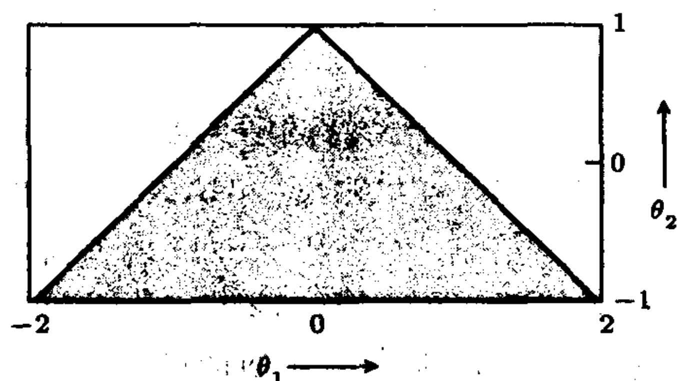
以上条件与 AR(2) 过程平稳性的必要条件 (3.2.18) 完全类似。

自相关函数　利用 (3.3.3)，该过程的方差为

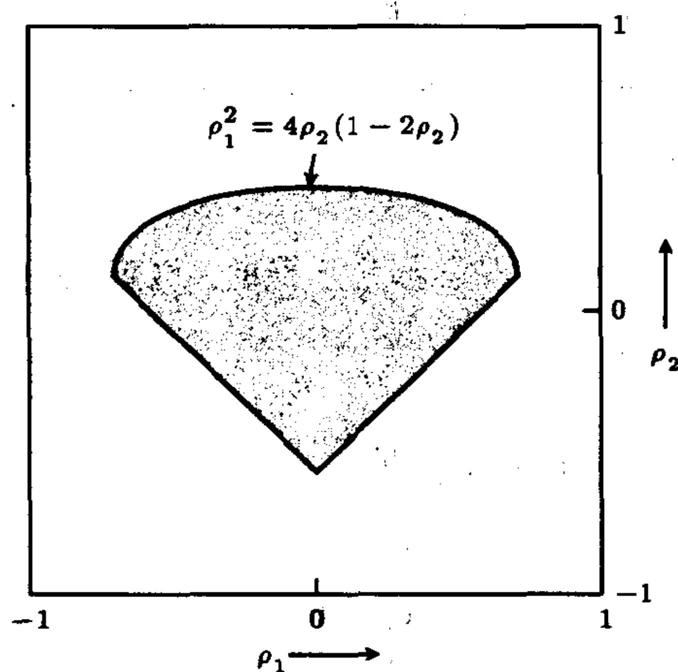
$$\gamma_0 = \sigma_a^2(1 + \theta_1^2 + \theta_2^2)$$

由 (3.3.4)，自相关函数是

$$\begin{aligned} \rho_1 &= \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2 &= \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \\ \rho_k &= 0 \qquad (k > 3) \end{aligned} \tag{3.3.11}$$



(a) θ_1, θ_2 的允许域



(b) ρ_1, ρ_2 的允许域

图 3.8 可逆 MA(2) 过程的允许域

由 (3.3.10) 和 (3.3.11) 可知，可逆 MA(2) 过程的第一、第二个自相关必定在下面曲线所界范围内：

$$\begin{aligned} \rho_2 + \rho_1 &= -0.5 \\ \rho_2 - \rho_1 &= -0.5 \\ \rho_1^2 &= 4\rho_2(1-2\rho_2) \end{aligned} \tag{3.3.12}$$

图 3.8(a) 给出了参数的可逆域 (3.3.10), 对应于 (3.3.12) 的自相关函数的区域在图 3.8(b) 中给出. 后一图表明，在给定区域不论怎

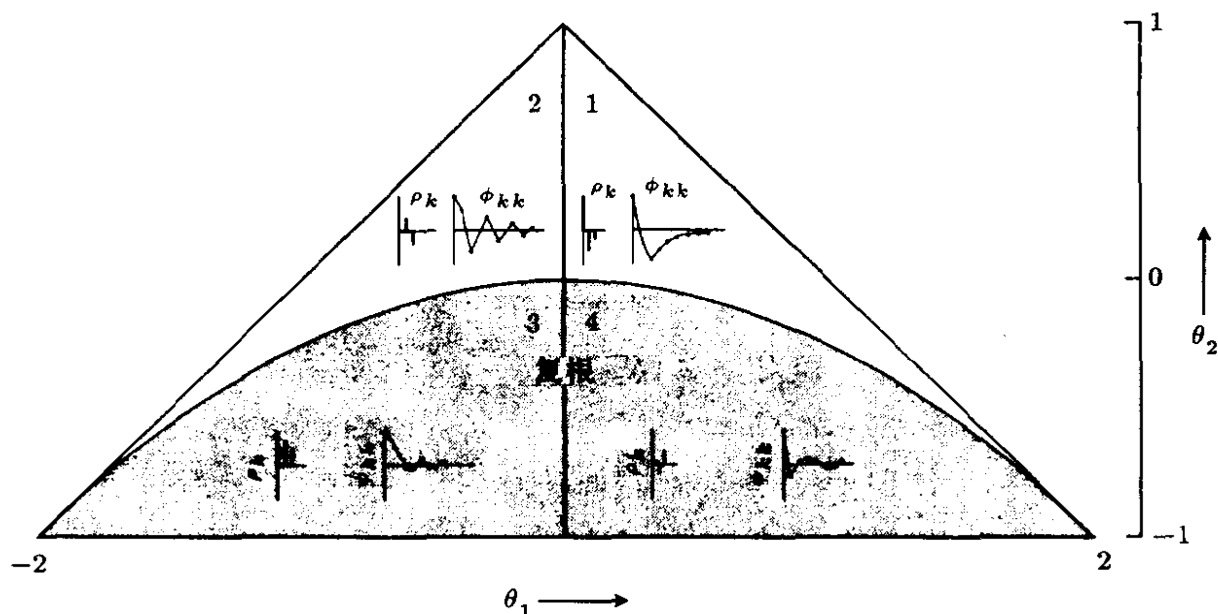


图 3.9 各种 MA(2) 模型的自相关、偏自相关函数 ρ_k 和 ϕ_{kk}(引自 [183])

样取一对自相关 ρ_1 和 ρ_2 都与模型为 MA(2) 过程的假定相容。如果自相关是相容的，那么参数 θ_1 和 θ_2,就可以通过求解非线性方程组 (3.3.11) 得出。为了便于计算，在第五部分"图表汇集"的图 C 中给出了曲线，当给定 ρ_1 和 ρ_2 时，可以直接读出 θ_1 和 θ_2 的值。

谱 由 (3.3.5) 得到，谱是

$$
\begin{aligned}
p(f) &= 2\sigma_a^2|1-\theta_1 e^{-i2\pi f}-\theta_2 e^{-i4\pi f}|^2 \\
&= 2\sigma_a^2[1+\theta_1^2+\theta_2^2-2\theta_1(1-\theta_2)\cos(2\pi f)-2\theta_2\cos(4\pi f)] \\
&\qquad (0<f<\frac{1}{2}) \qquad (3.3.13)
\end{aligned}
$$

如不计常数 $2\sigma_a^2$, 上式是二阶自回归过程谱的倒数。

偏自相关函数 MA(2) 过程偏自相关函数的严格表达式是很复杂的，但是可以得知，若特征方程 (3.3.9) 的根都是实根，则偏自相关函数被两项指数衰减之和所控制，若 (3.3.9) 具有实根，则被衰减的正弦波所控制。上述特性类似于 AR(2) 过程自相关的特性。图 3.9 给出了在可逆域中参数取不同值的自相关函数 (左边

曲线) 和偏自相关函数 (右边曲线) (该图引自 Stralkowski[183]) 。考查给出 AR(2) 过程自相关和对应偏相关的图 3.2, 并与图 3.9 对比，进一步表明了 AR(2) 和 MA(2) 过程之间的对偶关系。

3.3.5 自回归和滑动平均过程之间的对偶性

前面几节的结果说明了自回归和有限滑动平均过程之间存在对偶关系的特性，正如本章末尾的表 3.3 所阐明的，关于对偶性有下列结果：

1. 在一个 p 阶平稳自回归模型中， a_t 可表示为既往 $\tilde{z}$ 的有限加权和，换言之， $\tilde{z}_t$ 可表为既往 a 的无限加权和：

$$\tilde{z}_t = \phi^{-1}(B)a_t$$

同样，在一个 q 阶滑动平均模型中， $\tilde{z}_t$ 可表示为既往 a 的有限加权和，换言之， a_t 可表为既往 $\tilde{z}$ 的无限加权和：

$$\theta^{-1}(B)\tilde{z}_t = a_t$$

2. 有限的 MA 过程具有在某点之外全为零的自相关函数，但由于它等价于一个无限阶的 AR 过程，因此其偏自相关函数无限伸延，且被衰减指数和 (或) 衰减正弦波所控制。与此相反， AR 过程具有在某点之外全为零的偏自相关函数，但是它的自相关函数无限伸延，且由衰减指数和 (或) 衰减正弦波混合生成。

3. 对于一个有限 p 阶自回归过程，其参数不必满足任何条件就能保证可逆性，然而，为满足平稳性， $\phi(B)=0$ 的根必须都在单位圆外。与此相反， MA 过程的参数不需要满足任何条件就能保证平稳性，然而，为满足可逆性， $\theta(B)=0$ 的根必须都在单位圆外。

4. 滑动平均过程的谱与对应的自回归过程的谱存在互逆关系。

3.4 自回归滑动平均混合过程

3.4.1 平稳与可逆性质

在节 3.1.4 中我们已经注意到，为了使参数简约，在模型中可能必须既包含自回归项又包含滑动平均项。因此，可能需要使用混合的自回归和滑动平均模型

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \cdots \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \qquad (3.4.1)$$

也可写成

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\tilde{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t$$

或

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

其中 $\phi(B)$ 和 $\theta(B)$ 分别是 B 的 p 和 q 次多项式。

今后我们称这样过程为 ARMA(p,q) 过程。可以从下面两个方面来理解：

1. 将其看作是 p 阶自回归过程

$$\phi(B)\tilde{z}_t = e_t$$

而 e_t 为 q 阶滑动平均过程

$$e_t = \theta(B)a_t$$

2. 将其看作是 q 阶滑动平均过程

$$\tilde{z}_t = \theta(B)b_t$$

而 b_t 为 p 阶自回归过程

$$\phi(B)b_t = a_t$$

故有

$$\phi(B)\tilde{z}_t = \theta(B)\phi(B)b_t = \theta(B)a_t$$

显然，(3.4.1) 右边的滑动平均项不会影响节 3.2.1 中有关自回归过程平稳条件的讨论。若特征方程 $\phi(B)=0$ 的所有根都在单位圆外，那么，$\phi(B)\widetilde{z}_t=\theta(B)a_t$ 就定义一个平稳模型。与此类似，要使过程是可逆的，$\theta(B)=0$ 的根必须都在单位圆外。

因此，平稳可逆 ARMA(p,q) 过程 (3.4.1) 既有无限阶的滑动平均表示：

$$\widetilde{z}_t=\psi(B)a_t=\sum_{j=0}^{\infty}\psi_j a_{t-j}$$

这里 $\psi(B)=\phi^{-1}(B)\theta(B)$, 又有无限阶的自回归表示：

$$\pi(B)\widetilde{z}_t=\widetilde{z}_t-\sum_{j=1}^{\infty}\pi_j\widetilde{z}_{t-j}=a_t$$

这里，$\pi(B)=\theta^{-1}(B)\phi(B)$, 权 ψ_j 和 π_j 都是绝对可和的。权 ψ_j 由关系式 $\phi(B)\psi(B)=\theta(B)$ 确定，且满足

$$\psi_j=\phi_1\psi_{j-1}+\phi_2\psi_{j-2}+\cdots+\phi_p\psi_{j-p}-\theta_j \quad (j>0)$$

其中 $\psi_0=1,\psi_j=0$ 当 $j<0$, 及 $\theta_j=0$, 当 $j>q$. 另外，权 π_j 由关系式 $\theta(B)\pi(B)=\phi(B)$ 确定，并满足

$$\pi_j=\theta_1\pi_{j-1}+\theta_2\pi_{j-2}+\cdots+\theta_q\pi_{j-q}+\phi_j \quad (j>0)$$

这里，$\pi_0=-1,\pi_j=0$, 当 $j<0$, 及 $\phi_j=0$, 当 $j>p$. 由这些关系式和参数 ϕ_i 及 θ_i, 可以很容易地递推计算出权 ψ_j 和 π_j .

3.4.2 混合过程的自相关函数和谱

自相关函数　运用节 3.2.2 中对自回归过程使用的类似方法，混合过程的自相关过程可以推出。(3.4.1) 两边乘以 $\widetilde{z}_{t-k}$ 并取期望，我们看到自协方差函数满足差分方程

$$\gamma_k=\phi_1\gamma_{k-1}+\cdots+\phi_p\gamma_{k-p}+\gamma_{za}(l)-\theta_1\gamma_{za}(k-1)-\cdots-\theta_q\gamma_{za}(k-q)$$

这里 $\gamma_{za}(k)$ 是 z 和 a 的协方差函数，定义为 $\gamma_{za}(k) = E[\tilde{z}_{t-k}a_t]$。由于 $\tilde{z}_{t-k}$ 只依赖于 $t-k$ 时刻之前的“冲击”，这一点由其无限滑动平均表示 $\tilde{z}_{t-k} = \psi(B)a_{t-k} = \sum_{j=0}^{\infty}\psi_j a_{t-k-j}$ 可以看出，由此得到

$$\gamma_{za}(k) = \begin{cases} 0 & , \quad k > 0 \\ \psi_{-k}\sigma_a^2 & , \quad k < 0 \end{cases}$$

因此，前文关于 γ_k 的方程可表示为

$$\begin{aligned}\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} \\ - \sigma_a^2(\theta_k\psi_0 + \theta_{k-1}\psi_1 + \cdots + \theta_q\psi_{q-k})\end{aligned} \tag{3.4.2}$$

按照约定，取 $\theta_0 = -1$。

我们看到，(3.4.2) 蕴含

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} \qquad (k > q+1)$$

因而有

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \qquad (k > q+1) \tag{3.4.3}$$

或是

$$\phi(B) = \rho_k = 0 \qquad (k > q+1)$$

因此，由 (3.4.2) 看到，对于 ARMA(p,q) 过程而言，只有 q 个自相关 $\rho_q, \rho_{q-1}, \cdots, \rho_1$ 的值依赖于 q 个滑动平均参数 $\boldsymbol{\theta}$ 的选取，当然，这些自相关也依赖于 p 个自回归参数 $\boldsymbol{\phi}$。并且，p 个自相关 $\rho_q, \rho_{q-1}, \cdots, \rho_{q-p+1}$ 的值为差分方程 $\phi(B)\rho_k = 0, k > q+1$ 提供了必须的初值，高阶自相关全部可由差分方程确定。若 $q-p<0$，全体自相关函数 $\rho_j, j = 1, 2, \cdots$ 是由衰减指数和 (或) 衰减正弦波生成，其性质由多项式 $\phi(B)$ 和初值决定。然而，若 $q-p > 0$ 时，就会有 $q-p+1$ 个初值 $\rho_0, \rho_1, \cdots, \rho_{q-p}$ 不遵从一般的衰减变化形式。这一事实在识别混合序列时是有用的。

方差　当 $k=0$ 时有

$$\gamma_0=\phi_1\gamma_1+\cdots+\phi_p\gamma_p+\sigma_a^2(1-\theta_1\psi_1-\cdots-\theta_q\psi_q) \tag{3.4.4}$$

用此式及 $k=1,2,\cdots,p$ 的 p 个方程 (3.4.2), 我们可以解出 $\gamma_0,\gamma_1,\cdots,\gamma_p$。

谱　利用 (3.1.12), 混合过程的谱是

$$\begin{aligned}p(f)&=2\sigma_a^2\frac{|\theta(e^{-i2\pi f})|^2}{|\phi(e^{-i2\pi f})|^2}\\&=2\sigma_a^2\frac{|1-\theta_1e^{-i2\pi f}-\cdots-\theta_qe^{-i2\pi qf}|^2}{|1-\phi_1e^{-i2\pi f}-\cdots-\phi_qe^{-i2\pi pf}|^2}\\&\qquad(0<f<\frac{1}{2})\end{aligned} \tag{3.4.5}$$

偏自相关函数　过程 (3.4.1) 可以写成

$$a_t=\theta^{-1}(B)\phi(B)\widetilde{z}_t$$

这里 $\theta^{-1}(B)$ 是 B 的无穷级数, 因此混合过程的偏自相关函数是无限伸延的, 其特征就象纯滑动平均过程的偏自相关函数。换言之, 混合过程的偏自相关函数被复合的衰减指数和 (或) 衰减正弦波所控制, 衰减特性由滑动平均过程的阶数和具体参数决定。

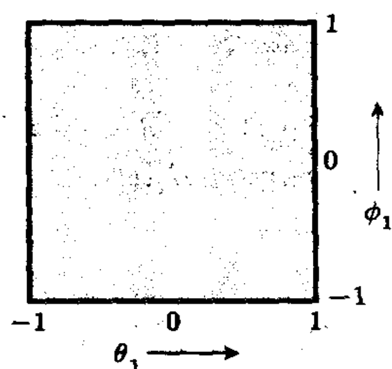
3.4.3 一阶自回归 ——一阶滑动平均过程

一类在实际中相当重要的混合过程是一阶自回归 ——一阶滑动平均的 ARMA(1,1) 过程

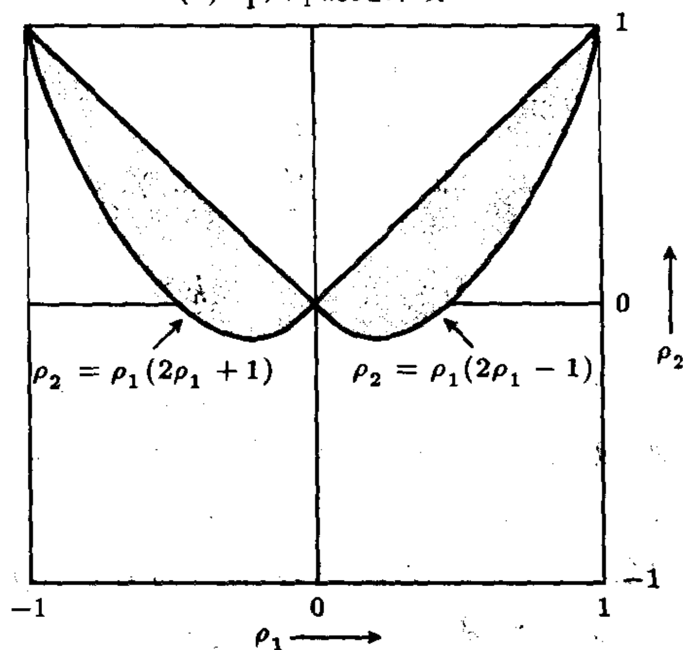
$$\widetilde{z}_t-\phi_1\widetilde{z}_{t-1}=a_t-\theta_1a_{t-1} \tag{3.4.6}$$

即

$$(1-\phi_1B)\widetilde{z}_t=(1-\theta_1B)a_t$$



(a) θ_1, ϕ_1 的允许域



(b) ρ_1, ρ_2 的允许域

图 3.10 平稳可逆 ARMA(1,1) 过程的允许域

我们现在来推导它的重要性质。

平稳性和可逆性条件 我们首先注意到，若 $-1 < \phi_1 < 1$ 过程为平稳，若 $-1 < \theta_1 < 1$ 过程是可逆，因此参数的允许空间是图 3.10(a) 中所示正方形。另外，对于平稳和可逆过程 ARMA(1,1)，由关系式 $\psi_1 = \phi_1\psi_0 - \theta_1 = \phi_1 - \theta_1$ 和 $\psi_j = \phi_1\psi_{j-1}, j > 1$ 我们求得权 ψ_j 由 $\psi_j = (\phi_1 - \theta_1)\phi_1^{j-1}, j > 1$ 给出，与此类似，易见 $\pi_j = (\phi_1 - \theta_1)\theta_1^{j-1}, j > 1$.

自相关函数 由 (3.4.2) 和 (3.4.4) 我们得到

$$\gamma_0 = \phi_1\gamma_1 + \sigma_a^2(1-\theta_1\psi_1)$$
$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2$$
$$\gamma_k = \phi_1\gamma_{k-1} \qquad (k \geqslant 2)$$

其中 $\psi_1 = \phi_1 - \theta_1$ 。因此前两方程对 γ_0 对 γ_1 求解，得到过程的自协方差函数为

$$\begin{aligned}\gamma_0 &= \frac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma_a^2 \\ \gamma_1 &= \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\sigma_a^2 \\ \gamma_k &= \phi_1\gamma_{k-1} \qquad (k \geqslant 2)\end{aligned} \tag{3.4.7}$$

从最后的方程可得 $\rho_k = \phi_1\rho_{k-1}, k \geqslant 2$, 即是 $\rho_k = \rho_1\phi_1^{k-1}, k > 1$ 。因此，自相关函数从初值 ρ_1 开始指数衰减，而 ρ_1 的值既依赖于 θ_1 也依赖于 ϕ_1[①] 。正如图 3.11 所示，若 ϕ_1 为正， ρ_k 的衰减是平滑的， ϕ_1 为负，则正负交替地衰减。另外， $(\phi_1-\theta_1)$ 决定了 ρ_k 的符号，并决定了衰减发生于正侧还是负侧。

由 (3.4.7) 式知，前两个自相关函数可用过程参数表示如下：

$$\begin{aligned}\rho_1 &= \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} \\ \rho_2 &= \phi_1\rho_1\end{aligned} \tag{3.4.8}$$

“图表汇集”中的图 C 构造出了方程 (3.4.8) 的解，若已知 ρ_1 和 ρ_2，可以从表中读出 ϕ_1 和 θ_1 。用估计值 r_1 和 r_2 代替 ρ_1 和 ρ_2，便可得到参数 ϕ_1 和 θ_1 的初估计。

利用 (3.4.8) 和平稳、可逆性条件可以证明， ρ_1, ρ_2 必定在下

[①] 与 AR(1) 过程不同， AR(1) 的自相关函数是从初值 $\rho_0=1$ 开始衰减。

面区域中：

$$\begin{aligned}&|\rho_2| < |\rho_1|\\&\rho_2 > \rho_1(2\rho_1+1) \qquad \rho_1 < 0\\&\rho_2 > \rho_1(2\rho_1-1) \qquad \rho_1 > 0\end{aligned} \tag{3.4.9}$$

图 3.10(b) 给出了 ρ_1 和 ρ_2 的允许范围 — 对混合的 (1,1) 平稳可逆过程而言，怎样的 ρ_1 和 ρ_2 才是可能的。

偏自相关函数　混合 ARMA(1,1) 过程 (3.4.6) 的偏自相关函数由一个简单的初值 $\phi_{11}=\rho_1$ 组成，由此可知，其特征与纯 MA(1) 过程偏自相关函数类似，也是被衰减指数所控制。因此，如图 3.11 中所示，当 θ_1 为正时，偏自相关函数被平滑的衰减指数所控制，衰减起始值为 ρ_1，符号由 $(\phi_1-\theta_1)$ 的符号决定。完全类似，当 θ_1 为负时，偏相关函数是被衰减指数所控制，振荡式地

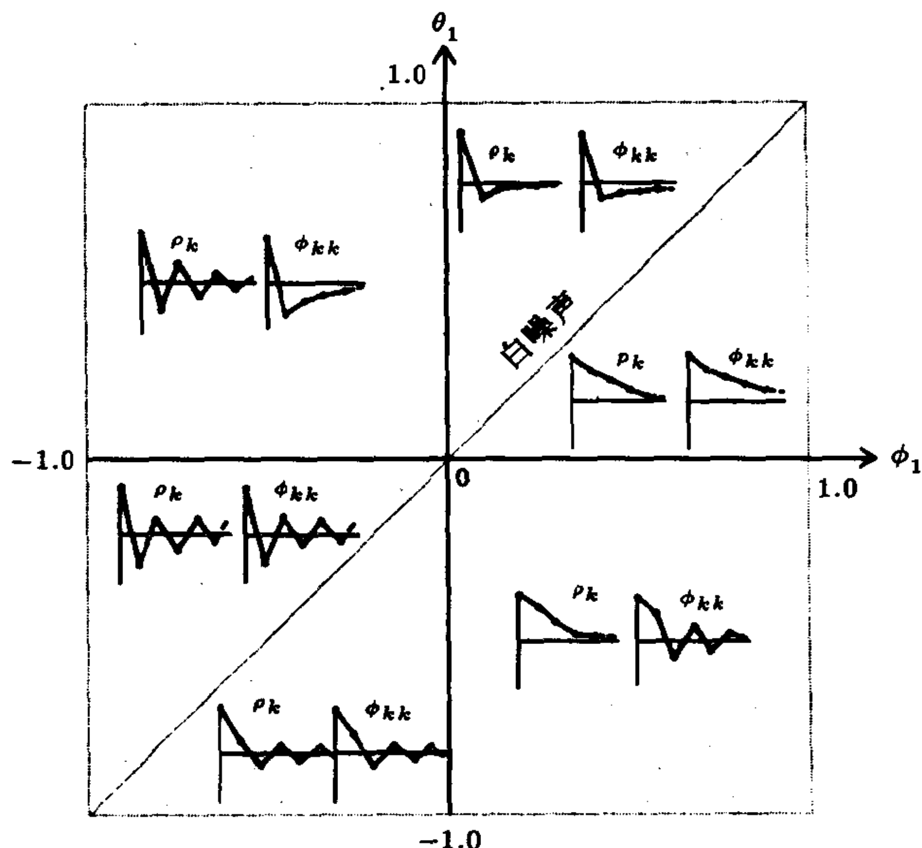


图 3.11　各类 ARMA(1,1) 模型的自相关 ρ_k 和偏自相关函数 ϕ_{kk}

表 3.2 对 ARMA(1,1) 模型 $(1-0.8)\widetilde{z}_t=(1+0.6B)a_t$, $\sigma_a^2=1$ 用 k 阶自回归"逼近"的自回归系数 ϕ_{kj} 和 $\sigma_k 2=r_0(1-\sum_{j=1}^{k}\phi_{kj}\rho_j)$ 的值

k	σ_k^2	"自回归系数"$\phi_{kj,j=1,\cdots,k}$						
1	1.304	0.893						
2	1.084	1.260	−0.411					
3	1.028	1.354	−0.697	0.227				
4	1.010	1.384	−0.709	0.407	−0.133			
5	1.003	1.394	−0.822	0.469	−0.242	0.079		
·	·	·	·	·	·	·		
·	·	·	·	·	·	·		
·	·	·	·	·	·	·		
∞	1.000	1.400	−0.840	0.504	−0.302	0.181	−0.109	···

* 对角线元素是偏自相关函数的值 ϕ_{kk} .

衰减，起始值为 ρ_1, 其符号由 $(\phi_1-\theta_1)$ 决定。

为了说明 ARMA(1,1) 过程偏自相关函数的特征及过程的其它特性，我考虑模型 $(1-0.8B)\widetilde{z}_t=(1+0.6B)a_t$, 这里 $\phi_1=0.8,\theta_1=-0.6$ 。由 (34.7) 和 (3.4.8), 我们求得过程的方差为 $\gamma_0=6.444$ 及 $\rho_1=0.893$ 。另外，自相关函数满足 $\rho_j=0.8\rho_{j-1},j>2$, 因此 $\rho_j=0.893(0.8)^{j-1},j>2$ 。通过求解与该过程相应的 Yule-Walker 方程可以得到自回归系数 $\phi_{k1},\cdots\phi_{kk}$ 。在表 3.2 中给出了 AR 的阶数为 $k=1$ 到 $k=5$ 的残差方差 $\sigma_k^2=\gamma_0(1-\sum_{i=1}^{k}\phi_{ki}\rho_i)$(即是 $\widetilde{z}_t-\sum_{i=1}^{k}\phi_{ki}\widetilde{z}_{t-i}$ 的方差)。注意，表 3.2 中对角线元素是该过程偏自相关函数的值 ϕ_{kk}, 由于 θ_1 的符号为负，故呈现振荡式指数衰减。另外，表中最后一行是该过程的 π_i 权，其值可由 $\pi_1=\phi_1-\theta_1=1.4$ 及 $\pi_j=\theta_1\pi_{j-1}=-0.6\pi_{j-1},j>1$ 得出，随着 k 增大，过程用 k 阶 AR"逼似"的系数值 ϕ_{kj} 趋近于 π_j 的值。最后我们注意到，当 $k>4$ 时，σ_k^2 的值非常接近于 σ_a^2, 这意味着用一个 4 阶 (或更高阶) 的自回归表示将能很好地逼近 ARMA(1,1) 模型，但这是以增加参数为代价的。

表 3.3 自回归、滑动平均及混合 ARMA 过程性质的总结

	自回归过程	滑动平均过程	混合过程
模型由过去的 $\tilde{z}$ 表示	$\phi(B)\tilde{z}_t=a_t$	$\theta^{-1}(B)\tilde{z}_t=a_t$	$\theta^{-1}(B)\phi(B)\tilde{z}_t=a_t$
模型由过去的 a 表示	$\tilde{z}_t=\phi^{-1}(B)a_t$	$\tilde{z}_t=\theta(B)z_t$	$\tilde{z}_t=\phi^{-1}(B)\theta(B)a_t$
权数 π	有限序列	无限序列	无限序列
权数 ψ	无限序列	有限序列	无限序列
平稳性条件	$\phi(B)=0$ 的根在单位圆外	无条件平稳	$\phi(B)=0$ 的根在单位圆外
可逆性条件	无条件可逆	$\theta(B)=0$ 的根在单位圆外	$\theta(B)=0$ 的根在单位圆外
自相关函数	无限项 (指数衰减和(或) 正弦振荡衰减) 拖尾	有限项 q 步后截尾	无限项 (指数衰减和(或) 正弦振荡衰减) 拖尾
偏自相关函数	有限项 p 步后截尾	无限项 (指数衰减和(或) 正弦振荡衰减) 拖尾	无限项 ($p-q$ 步后指数衰减和 (或) 正弦振荡衰减) 拖尾

3.4.4 总结

图 3.12 集中了平稳且可逆过程 AR(2),MA(2), 及 ARMA(1,1) 的参数和自相关 ρ_1,ρ_1 的允许域。表 3.3 总结了混合自回归滑动平均过程的性质，并归纳了有关自回归、滑动平均及混合过程的重要结果，所有这些在第 6 章中对观测序列作模型识别是必不可少的。在第 4 章中我们将对混合 ARMA 模型加以推广，使之能够描述实际中经常遇到的一类非平稳特性。

A3.1 一般线性过程的自协方差函数，自协方差生成函数及平稳性条件

自协方差函数 线性过程

$$\tilde{z}_t=\sum_{j=0}^{\infty}\psi_j a_{t-j}$$

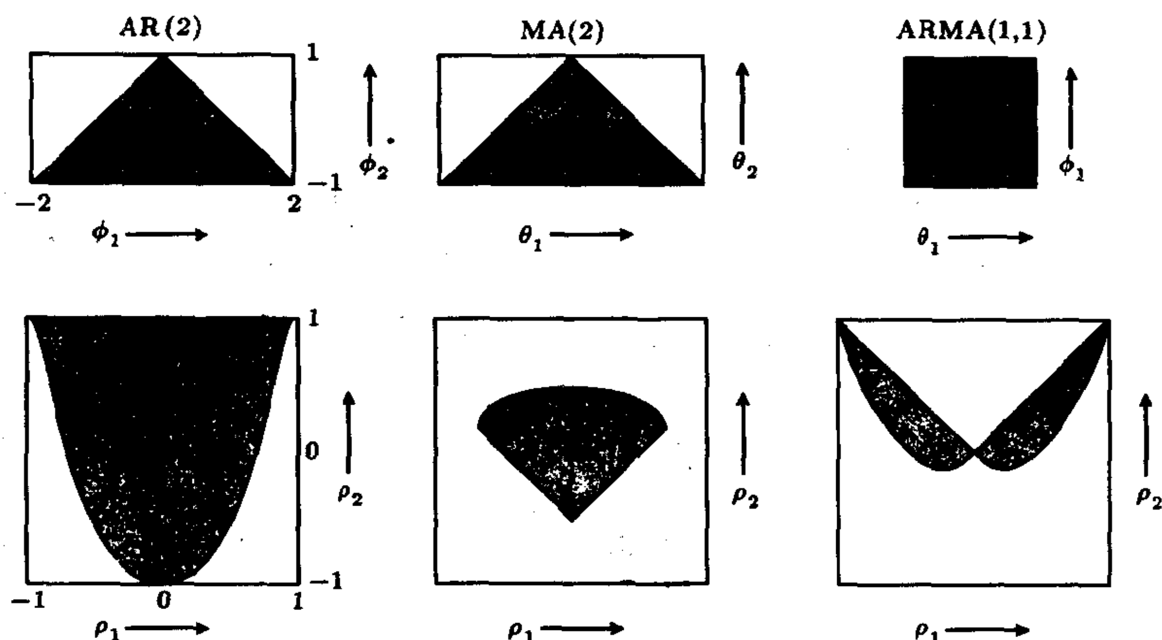


图 3.12 在平稳可逆条件下，过程 AR(2),MA(2),ARMA(1,1) 的参数及 ρ_1,ρ_2 的允许域

其中， $\psi_0 = 1$, 该过程的 k 步滞后自协方差函数显然为

$$
\begin{aligned}
\gamma_k = E[\widetilde{z}_t \widetilde{z}_{t+k}] &= E\left[\sum_{j=0}^{\infty}\sum_{h=0}^{\infty}\psi_j\psi_h a_{t-j}a_{t+k-h}\right] \\
&= \sigma_a^2 \sum_{j=0}^{\infty}\psi_j\psi_{j+k}
\end{aligned}
\tag{A3.1.1}
$$

这里，利用了白噪声自协方差函数的性质 (3.1.2) 。

自协方差生成函数　自协方差生成函数为

$$
\gamma(B) = \sum_{k=-\infty}^{\infty}\gamma_k B^k \tag{A3.1.2}
$$

将前文得出的 (A3.1.1) 代入后得到

$$
\begin{aligned}
\gamma(B) &= \sigma_a^2 \sum_{k=-\infty}^{\infty}\sum_{j=0}^{\infty}\psi_j\psi_{j+k}B^k \\
&= \sigma_a^2 \sum_{j=0}^{\infty}\sum_{k=-j}^{\infty}\psi_j\psi_{j+k}B^k
\end{aligned}
$$

由于 $\psi_k = 0$, 当 $h < 0$ 时，记 $j + k = h$, 从而有 $k = h - j$, 于是我们得到

$$
\begin{aligned}
\gamma(B) &= \sigma_a^2 \sum_{j=0}^{\infty}\sum_{h=0}^{\infty}\psi_j\psi_h B^{h-j} \\
&= \sigma_a^2 \sum_{h=0}^{\infty}\psi_h B^h \sum_{j=0}^{\infty}\psi_j B^{-j}
\end{aligned}
$$

因此

$$
\gamma(B) = \sigma_a^2\psi(B)\psi(B^{-1}) = \sigma_a^2\psi(B)\psi(F) \tag{A3.1.3}
$$

这就是在正文中得出 (3.1.11) 时所引用的结果.

平稳性条件　若将 $B = e^{-i2\pi f}$ 和 $F = B^{-1} = e^{i2\pi f}$ 代入自协方差生成函数 (A3.1.2), 可得半功率谱, 因此线性过程的功率谱为

$$\begin{aligned} p(f) &= 2\sigma_a^2 \psi(e^{-i2\pi f})\psi(e^{i2\pi f}) \\ &= 2\sigma_a^2 |\psi(e^{-i2\pi f})|^2 \qquad (0 < f < \frac{1}{2}) \end{aligned} \tag{A3.1.4}$$

由此得到过程的方差

$$\sigma_a^2 = \int_0^{1/2} p(f)df = 2\sigma_a^2 \int_0^{1/2} \psi(e^{-i2\pi f})\psi(e^{i2\pi f})df \tag{A3.1.5}$$

若积分 (A3.1.5) 收敛, 可以证明 [98] 无穷级数 $\psi(B)$ 必定在单位圆内或圆上收敛. 立刻得知, 对于线性过程 $\tilde{z}_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, 系数 ψ_j 的绝对可和性条件: $\sum_{j=0}^{\infty} |\psi_j| < \infty$ 蕴含 (见 [63],[93]) 无穷和 $\sum_{j=0}^{\infty} \psi_j a_{t-j}$ 概率为 1 地收敛, 因此它表述一个有效的平稳过程.

A3.2　计算自回归参数估计值的递推方法

假设对某个时间序列用 AR(p) 模型进行拟合, 参数估计值已经得到, 现在我们来说明, 如何在此基础上得出对同一序列拟合 AR($p+1$) 模型并得出参数的 Yule-Walker 估计. 如 3.2.6 节所述, 这种递推算法也可以用来近似计算偏相关函数.

为了说明递推算法, 考虑方程 (3.2.35) . 对于 $k = 2, 3$,Yule-Walker 估计可由下式得到

$$\begin{aligned} r_2 &= \hat{\phi}_{21} r_1 + \hat{\phi}_{22} \\ r_1 &= \hat{\phi}_{21} + \hat{\phi}_{22} r_1 \end{aligned} \tag{A3.2.1}$$

及

$$\begin{aligned} r_3 &= \hat{\phi}_{31} r_2 + \hat{\phi}_{32} r_1 + \hat{\phi}_{33} \\ r_2 &= \hat{\phi}_{31} r_1 + \hat{\phi}_{32} + \hat{\phi}_{33} r_1 \\ r_1 &= \hat{\phi}_{31} + \hat{\phi}_{32} r_1 + \hat{\phi}_{33} r_2 \end{aligned} \tag{A3.2.2}$$

利用 (A3.2.2) 中最后两个方程，系数 $\hat{\phi}_{31}$ 和 $\hat{\phi}_{32}$ 可由 $\hat{\phi}_{33}$ 表示。解可用矩阵写成如下形式：

$$\begin{pmatrix} \hat{\phi}_{31} \\ \hat{\phi}_{32} \end{pmatrix} = \boldsymbol{R}_2^{-1} \begin{pmatrix} r_2 - \hat{\phi}_{33} r_1 \\ r_1 - \hat{\phi}_{33} r_2 \end{pmatrix} \tag{A3.2.3}$$

这里

$$\boldsymbol{R}_2 = \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}$$

现可将 (A3.2.3) 写成

$$\begin{bmatrix} \hat{\phi}_{31} \\ \hat{\phi}_{32} \end{bmatrix} = [\boldsymbol{R}_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix} - \hat{\phi}_{33} \boldsymbol{R}_2^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{A3.2.4}$$

(A3.2.1) 可以改写为

$$\begin{bmatrix} \hat{\phi}_{21} \\ \hat{\phi}_{22} \end{bmatrix} = \boldsymbol{R}_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix}$$

利用这一事实得到，(A3.2.4) 化为

$$\begin{bmatrix} \hat{\phi}_{31} \\ \hat{\phi}_{32} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{21} \\ \hat{\phi}_{22} \end{bmatrix} - \hat{\phi}_{33} \begin{bmatrix} \hat{\phi}_{22} \\ \hat{\phi}_{21} \end{bmatrix}$$

也即

$$\begin{aligned} \hat{\phi}_{31} &= \hat{\phi}_{21} - \hat{\phi}_{33} \hat{\phi}_{22} \\ \hat{\phi}_{32} &= \hat{\phi}_{22} - \hat{\phi}_{33} \hat{\phi}_{22} \end{aligned} \tag{A3.2.5}$$

为了完成 $\hat{\phi}_{31}$ 和 $\hat{\phi}_{32}$ 的计算，我们需要 $\hat{\phi}_{33}$ 的表达式。将 (A3.2.5) 代入方程 (A3.2.2) 我们得到

$$\hat{\phi}_{33} = \frac{r_3 - \hat{\phi}_{21} r_2 - \hat{\phi}_{22} r_1}{1 - \hat{\phi}_{21} r_1 - \hat{\phi}_{22} r_2} \tag{A3.2.6}$$

因此偏自相关 $\hat{\phi}_{33}$ 最先由 $\hat{\phi}_{21}$ 和 $\hat{\phi}_{22}$ 算出，利用 (A3.2.6)，另两个系数 $\hat{\phi}_{31}$ 和 $\hat{\phi}_{32}$ 也可由 (A3.2.5) 的出．对于一般情形，由 Durbin[87] 给出了下面的递推公式：

$$\hat{\phi}_{p+1,j} = \hat{\phi}_{pj} - \hat{\phi}_{p+1,p+1}\hat{\phi}_{p,p-j+1} \qquad (j = 1, 2, \cdots, p) \tag{A3.2.7}$$

$$\hat{\phi}_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^{p}\hat{\phi}_{pj}r_{p+1-j}}{1 - \sum_{j=1}^{p}\hat{\phi}_{pj}r_j} \tag{A3.2.8}$$

例：为说明起见，我们用一个 AR(3) 模型去拟合 Wölfer 太阳黑子数，对相应模型的参数 $\hat{\phi}_{31}, \hat{\phi}_{32}$ 和 $\hat{\phi}_{33}$ 进行计算．精确到第三位小数的自相关估计是 $r_1 = 0.806, r_2 = 0.428$ 及 $r_3 = 0.070$ ．于是

$$\hat{\phi}_{21} = \frac{r_1(1 - r_2)}{1 - r_1^2} = 1.316$$

$$\hat{\phi}_{22} = \frac{r_2 - r_1^2}{1 - r_1^2} = -0.632$$

利用 (A3.2.6)，我们得到

$$\hat{\phi}_{33} = \frac{0.070 - (1.316)(0.428) + (0.632)(0.806)}{1 - (1.316)(0.806) + (0.632)(0.428)} = 0.077$$

在 (A3.2.5) 中将 $\hat{\phi}_{21}, \hat{\phi}_{22}$ 和 $\hat{\phi}_{33}$ 的值代入，得出

$$\hat{\phi}_{31} = 1.316 + (0.077)(0.632) = 1.365$$

$$\hat{\phi}_{32} = -0.632 - (0.077)(1.316) = -0.733$$

在这里我们提请读者注意，递推算法得到的估计在某种程度上不同于逐次提高阶数拟合回归模型而得出的极大似然估计，递推算法对舍入误差非常灵敏，特别当过程接近于非平稳时，更是如此．

4

线性非平稳模型

许多实际中的时间序列 (例如，股票) 在演化过程中并不是具有固定不变的均值，尽管如此，除去局部水平 (或除去局部水平和趋势) 不同以外，序列显示出具有某种意义上的同质性，即序列的某一部分与任何其它部分极其相似. 我们假定对过程作适当的差分便可使之平稳化，这样就可以得到描述同质非平稳特性的模型. 我们现在考虑一类重要模型的性质，即相应序列经 d 阶差分后就成为平稳的混合自回归 — 滑动平均过程，我们称这类模型为求和自回归滑动平均 (ARIMA) 过程。

4.1 求和自回归滑动平均过程

4.1.1 非平稳一阶自回归过程

图 4.1 给出了在实际中遇到的四个时间序列，它们是在预测或控制问题中出现的，且都表现出前文提及的非平稳特性。序列 A,C, 和 D 表示取自于三个不同化工过程 (分别为浓度、温度和粘滞度) 的"非控制"输出. 我们收集这些序列是为了说明非控制、非量测干扰 (例如物料和周围温度的变化等) 对输出的影响. 序列 C 是通过先断开有关试验设备的控制器，随后记录下温度的变化而得到的. 序列 A 和 D 都是对原生产过程采样的，其间必须使输出产品的质量特征尽可能保持某一固定水平. 为了实现这种控

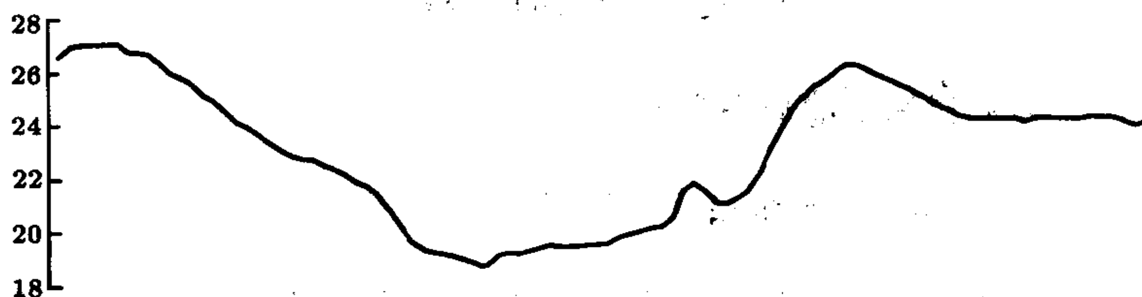
制，必须对其它变量进行操作，以此抵消输出量的变化．然而，这些操作对输出量的影响都是准确知道的，因此可以对控制措施进行数值补偿．总之，我们能够由观测序列计算出相应的值，这些值非常接近未采取修正措施时的观测．记录这些补偿值，并称



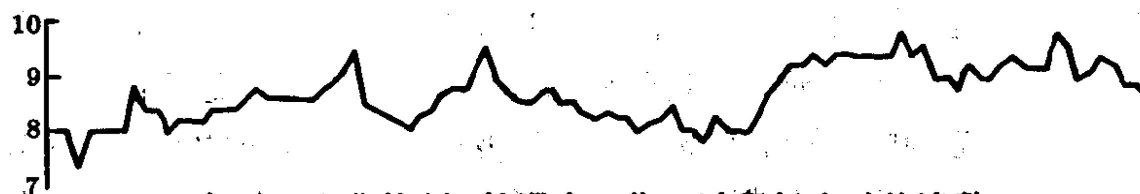
序列 A: “非控制”浓度，化工过程 2 小时的读数



序列 B: IBM 日股票价格



序列 C: “非控制”温度，化工过程每分钟的读数



序列 D: “非控制”粘滞度，化工过程每小时的读数

20 40 60 80 100

图 4.1 测和控制问题中出现的典型时间序列

表 4.1　平稳一阶自回归过程的前 11 个值

t	a_t	$\widetilde{z}_t$	t	a_t	$\widetilde{z}_t$
0		0.7	6	–0.8	22.8
1	0.1	1.5	7	0.8	46.4
2	–1.1	1.9	8	0.1	92.9
3	0.2	4.0	9	0.1	185.9
4	–2.0	6.0	10	–0.9	370.9
5	–0.2	11.8			

之为“非控制序列”. 序列 B 是 1961 年 5 月初一段时间的 IBM 日股票价格. 本书末的“时间序列汇集”中给出了前述各个序列的完整的清单. 在图 4.1 中给出了每个序列中 100 个顺序观测值的图, 各点之间用直线连接. 过程表现出的非平稳特性不会只有有限多种形式, 然而, 我们经常需要分析的经济和生产序列都表现出一种特殊的同质非平稳特性, 都能用对 ARMA 加以改进的随机模型来描述. 在第 3 章中已考虑了混合自回归滑动平均模型:

$$\phi(B)\widetilde{z}_t = \theta(B)a_t \tag{4.1.1}$$

这里, $\phi(B)$ 和 $\theta(B)$ 分别是 B 的 p 次及 q 次多项式. 为了保证平稳性, $\phi(B)=0$ 的根必须在单位圆外. 为得到非平稳过程, 一个很自然的方法就是放宽这种限制.

为了深入地了解这种可能性, 我们考虑一阶自回归模型:

$$(1-\phi B)\widetilde{z}_t = a_t \tag{4.1.2}$$

当 $|\phi|<1$ 时, 模型是平稳的. 让我们来研究 $\phi=2$ 时该过程的特性, 这时参数取值在平稳域外.

表 4.1 给出了正态单位随机偏差 a_t 和相应的序列值 $\widetilde{z}_t$,$\widetilde{z}_t$ 是按模型 $\widetilde{z}_t = 2\widetilde{z}_{t-1} + a_t$ 生成, 其中 $\widetilde{z}_0 = 0.7$. 图 4.2 是序列的散点图. 从图中可以看到, 经过短暂的过渡后, 序列就“脱轨”了, 实质上遵从指数规律, a 的作用对未来部分几乎不产生影响. 破坏平稳性的高阶过程所生成的序列, 其特性也完全类似. 并且, 不论模型右边是否引入滑动平均项, 这种特性本质上是相同的.

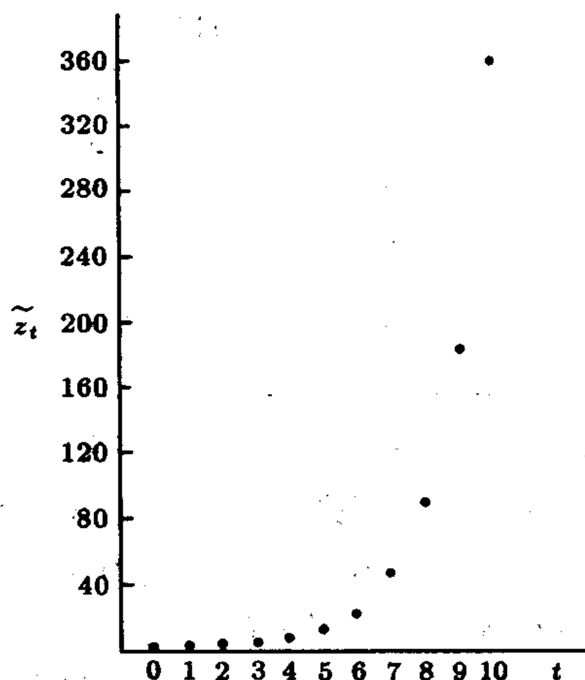


图 4.2 非平稳一阶自回归过程 $\widetilde{z}_t=2\widetilde{z}_t+a_t(\sigma_a^2=1)$ 的实现

4.1.2 描述同质性非平稳过程的一般模型

求和自回归滑动平均模型 尽管前文所述非平稳模型对于描述爆炸或进化特性(如细菌繁衍)是有价值的，但在本书中我们所讨论的并不是这种类情形. 到目前为止我们已经看到，若 $\phi(B)=0$ 的根在单位圆外，则 ARMA 是平稳的；若根在单位圆内，则过程呈现爆炸性的非平稳特性. 对我们来说，可供选择的唯一情形就是 $\phi(B)=0$ 的根在单位圆上了，由此而得出的模型对于描述同质非平稳时间序列是极其重要的. 特别，具有一个或多个单位根的模型往往能对非季节序列作很好的描述，在本章中我们将考虑这一问题[①].

我们来考虑模型

$$\varphi(B)\widetilde{z}_t=\theta(B)a_t \tag{4.1.3}$$

① 在第 9 章中，我们将考虑能构描述具有周期为 s 的季节性模型，其特征方程在单位圆上有 s 重根.

这里，$\varphi(B)$ 是非平稳自回归算子，$\varphi(B)=0$ 有 d 个单位根，其余的根都在单位圆外。于是我们可以将模型 (4.1.3) 表示成如下形式

$$\varphi(B)\widetilde{z}_t = \phi(B)(1-B)^d\widetilde{z}_t = \theta(B)z_t \tag{4.1.4}$$

式中 $\phi(B)$ 是平稳自回归算子。由于 $\nabla^d\widetilde{z}_t = \nabla^d z_t$, 对于 $d > 1, \nabla = 1-B$ 是差分算子，我们可将模型写为

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{4.1.5}$$

等价地，过程可以由下面两个方程定义：

$$\phi(B)w_t = \theta(B)a_t \tag{4.1.6}$$

及

$$w_t = \nabla^d z_t \tag{4.1.7}$$

因此我们看到，上述模型相当于假定序列经 d 阶差分后能表示为平稳可逆的 ARMA 过程。另一方面，当 $d > 1$ 时将 (4.1.7) 逆转得出

$$z_t = S^d z_t \tag{4.1.8}$$

其中 S 是由下式定义的无穷和算子：

$$\begin{aligned} Sx_t = \sum_{h=-\infty}^{t} x_h &= (1+B+B^2+\cdots)x_t \\ &= (1-B)^{-1}x_t = \nabla^{-1}x_t \end{aligned}$$

故

$$S = (1-B)^{-1} = \nabla^{-1}$$

算子 S^2 可类似定义为

$$\begin{aligned} S^2x_t &= Sx_t + Sx_{t-1} + Sx_{t-2} + \cdots \\ &= \sum_{i=-\infty}^{t}\sum_{h=-\infty}^{i} x_h = (1+2B+3B^2+\cdots)x_t \end{aligned}$$

对高阶 d 可以定义同样的运算。方程 (4.1.8) 表示：过程 (4.1.5) 可以通过对平稳过程 (4.1.6) 作 d 重求和 (或积分) 而得到，因此，我们称过程 (4.1.5) 为求和自回归滑动平均 (ARIMA) 过程。关于非平稳时间序列的 ARIMA 模型， Yaglom[209] 早就考虑过，这种模型对于预测和控制方面的问题是非常重要的，可参见 [42],[43],[44],[45],[46],[47],[49] 。为了进一步讨论非平稳过程，读者还可参见 Zadeh 和 Ragazzini[213],Kalman[124] 以及 Kalman 和 Bucy[125] 的有关文献。早期时间序列分析的处理办法是采用“变差方法”(见 Tintner[187],Rao 和 Tintner[165] 。) 但是，这种处理办法的动机、方法和目的都与这里讨论的截然不同。

从技术上说，无穷和算子 $S=(1-B)^{-1}$ 不能真正用来定义非平稳 ARIMA 过程，这是因为涉及到的无穷和是不收敛的。我们用有穷和算子取而代之，对任意正整数 m, 有穷和算子 S_m 由下式给出

$$S_m=(1+B+B^2+\cdots+B^{m-1})=\frac{1-B^m}{1-B}$$

类似地，两重的有穷和算子可定义为

$$\begin{aligned}S_m^{(2)}&=\sum_{j=0}^{m-1}\sum_{i=j}^{m-1}B^i=(1+2B+3B^2+\cdots+mB^{m-1})\\&=\frac{1-B^m-mB^m(1-B)}{(1-B)^2}\end{aligned}$$

这是因为 $(1-B)S_m^{(2)}=S_m-mB^m$, 等等。因此，一个求和 ARMA 过程 (例如 $d=1$) 与对应的平稳 ARMA 过程 $w_t=(1-B)z_t$ 之间的关系，可以表示为用直到过去的某一时间原点 $k(k<t)$ 之前的 w 来表示，即可表为

$$z_t=\frac{S_{t-k}}{1-B^{t-k}}w_t=\frac{1}{1-B^{t-k}}(w_t+w_{t-1}+\cdots+w_{k+1})$$

故有， $z_t = w_t + w_{t-1} + \cdots + w_{k+1} + z_k$ 可以看成是平稳过程 w 的有穷项之和再加上过程 z 在时刻 k 的初始值. 因此，对于由 (4.1.3) 生成的非平稳 ARIMA 过程，我们形式上确定其随机性质时，特别要说明该过程在某个有穷的过去时点 (当然可能是久远的过去)k 的初值. 然而，初始条件本身对于过程的的主要特征影响甚微，对初始条件的有关细节在本书中一般不予强调.

如第 1 章所述，模型 (4.1.5) 等价于将过程 z_t 表示为一个线性滤波器的输出 (仅当 $d=0$ 时为稳定的滤波), 相应的输入为白噪声 a_t . 换言之，我们可以认为该模型将相互依赖、非平稳的过程 z_t 转化成不相关的随机变量序列 a_t, 也就是将过程转化为白噪声.

若在 (4.1.5) 中，自回归算子 $\phi(B)$ 是 p 阶，差分次数取为 d, 滑动平均算子 $\theta(B)$ 是 q 阶，我们便称之为 (p,d,q) 阶的 ARIMA 模型，或简称 ARIMA(p,d,q) 过程.

ARIMA 模型的两种解释　我们先来说明， (4.1.5) 用于描述实际中的时间序列，是一种直观而合理的模型. 首先我们注意到，当 $|\phi|<1$ 或 $|\phi|>1$ 时，由一阶自回归过程 (4.1.2) 产生的序列，其局部特性很大程度依赖于 $\tilde{z}_t$ 的水平. 这和图 4.1 中那些序列的特性有所不同，图 4.1 中序列表现出其局部特性与自身水平是无关的.

如果我们要采用一种模型，以使过程的特性与其水平无关，那么必须选择自回归算子 $\varphi(B)$ 使得

$$\varphi(B)(\tilde{z}_t + c) = \varphi(B)\tilde{z}_t$$

这里， c 是任意常数. $\varphi(B)$ 必然有如下表式:

$$\varphi(B) = \phi_1(B)(1-B) = \phi_1(B)\nabla$$

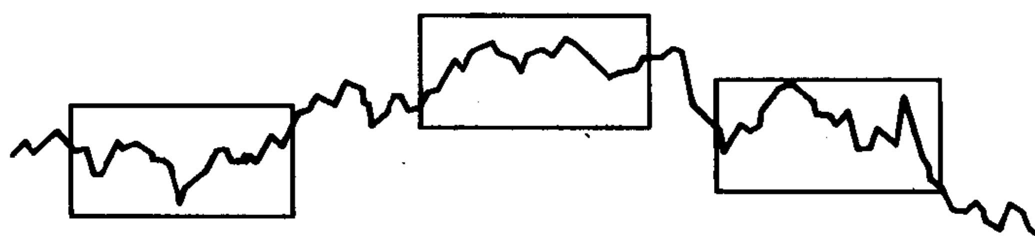
因此，具有前述性质的过程类可表示为

$$\phi_1(B)w_t = \theta(B)a_t$$

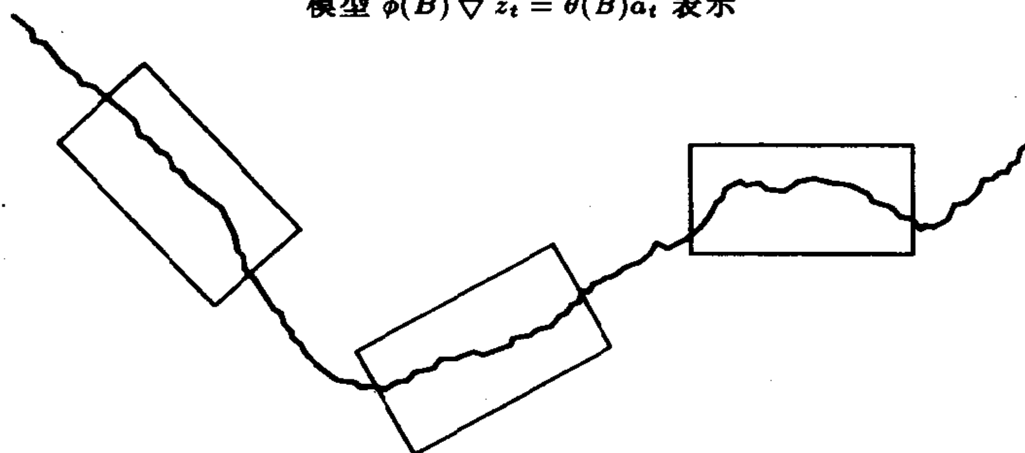
其中 $w_t = \nabla\tilde{z}_t = \nabla z_t$. 我们所要求的同质性排除了 w_t 呈爆炸性增长的可能. 因为这意味着: 或者 $\phi_1(B)$ 就是平稳自回归算子, 或者 $\phi_1(B) = \phi_2(B)(1-B)$, 使得 $\phi_2(B)w_t = \theta(B)a_t$, 这里 $w_t = \nabla^2 z_t$. 在后一种情况, 同样的讨论又可用于二阶差分, 如此等等.

最后, 我们得到的结论是: 为了表示非平稳而又具有同质性的时间序列, (4.1.3) 左边的算子应该形如 $\phi(B)\nabla^d$, 其中 $\phi(B)$ 是平稳自回归算子. 因此, 我们又要回到模型 (4.1.5) 。

为了从不同的角度来探讨这一模型, 我们考虑 (4.1.4) 中 $d = 0$ 的情形, 这时 $\phi(B)\tilde{z}_t = \theta(B)a_t$. 要求 $\phi(B)$ 的零点都在单位圆外, 这不仅保证了 $\tilde{z}_t$ 是零均值平稳过程, 也保证了 $\nabla z_t, \nabla^2 z_t, \nabla^3 z_t \cdots$ 都是零均值平稳的. 图 4.3(a) 给出了一种非平稳序列, 我们希望用前面的模型来表示. 除了水平值不同, 该序列是同质的, 也就



(a) 表现水平非平稳性的时间序列, 可以用模型 $\phi(B)\nabla z_t = \theta(B)a_t$ 表示



(b) 表现水平和斜率非平稳性的时间序列, 可以用模型 $\phi(B)\nabla^2 z_t = \theta(B)a_t$ 表示

图 4.3 两类同质非平稳特性

是说，序列的一部分和其它部分非常相似，只是在垂直方向上位置不同。我们不管其水平取何值，只要求每一段在差分之后是平稳零均值的，这样就能描述这种特性。我们采用下面的模型

$$\phi(B)\nabla z_t = \theta(B)a_t$$

图 4.3(b) 给出了经常出现的第二类平稳序列。这类序列既没有固定的水平，也没有固定的斜率，但如果我们容许这些特征不同，那么序列的特性仍是同质的。这种特性用模型表示便是

$$\phi(B)\nabla^2 z_t = \theta(B)a_t$$

该模型保证了在一阶、二阶之后的所有差分都是零均值平稳的，但是允许水平和斜率“任意”。

4.1.3 求和自回归滑动平均过程的一般形式

基于下文将提到的理由，我们要考虑 ARIMA 模型 (4.1.5) 稍加推广的形式，即增加一个常数项 θ_0，在有些场合这样做是有用的。我们将要用来描述时间序列较为一般的模型是下面的求和自回归滑动平均过程

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)z_t \tag{4.1.9}$$

其中

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

在下文中假定：

1. $\phi(B)$ 称为自回归算子，假设它是平稳的，即 $\phi(B) = 0$ 的根在单位圆外。

2. $\varphi(B) = \phi(B)\nabla^d$ 称为广义自回归算子，它是非平稳的。$\varphi(B) = 0$ 有 d 个根等于 1.

3. $\theta(B)$ 称为滑动平均算子，假设它是可逆的，即 $\theta(B)=0$ 的根在单位圆外.

当 $d=0$ 时，模型 (4.1.9) 表示一个平稳过程. 有关平稳性和可逆性的要求是独立使用的，一般说来，算子 $\phi(B)$ 和 $\theta(B)$ 的阶数是不同的. 在第 3 章中，我们已对 $p=1,2$ 的简单情形给出了平稳域的例子，同样也给出了 $q=1,2$ 的可逆域.

随机趋势和确定性趋势　在节 4.1.2 中我们已经看到，当略去常数项 θ_0 时，模型 (4.1.9) 能够表示具有随机趋势的序列，典型的例子是序列的水平和斜率都是随机变化的. 然而，通常我们还希望在模型中含有一个时间的确定性函数 $f(t)$，若允许 θ_0 非零，很自然地会考虑到确定性的 d 阶多项式趋势. 例如，当 $d=1$ 时，我们可以利用 $\theta_0\neq 0$ 的模型从非平稳噪声中去估计可能存在的确定性线性趋势. 由 (3.1.22) 知，允许 θ_0 非零，相当于容许

$$E[w_t]=E[\nabla^d z_t]=\mu_w=\frac{\theta_0}{1-\phi_1-\phi_2-\cdots-\phi_p}$$

为非零. 表示一般模型 (4.1.9) 的另一种办法，就是用 $\widetilde{w}_t=w_t-\mu_w$ 将其表示成平稳可逆 ARMA 过程的形式，即

$$\phi(B)\widetilde{w}_t=\theta(B)a_t \tag{4.1.10}$$

在实际应用中，如果没有确实的理由表明确定性分量存在，可以假设 w 的均值为零，除非经检验表明该假设与数据所提供的事实相矛盾. 显然，在许多应用中，随机性假设往往比确定性假设更现实一些. 这一点在时间序列的预测中特别重要，因为随机性趋势不要求序列依旧遵循过去发展的模式. 因此，当 $d>0$ 时，通常我们假设 $\mu_w=0$(或等价地，$\theta_0=0$)，只有当数据或问题的本身的性质非常显著时，才需要假设非零均值，或更一般地假设已知形式的确定性分量.

ARIMA 模型一些重要的特殊情形　在第 3 章中，我们已经熟悉了模型 (4.1.9) 的一些重要的特殊情形，即对应于 $d=0$ 的平

稳情形. 下述模型描述了一些经常出现的非平稳 ($d > 1$) 的特殊情形.

1. (0,1,1) 过程:

$$\begin{aligned}\nabla z_t &= a_t - \theta_1 a_{t-1}\\ &= (1-\theta_1 B)a_t\end{aligned}$$

对应于 $p=0, d=1, q=1, \phi(B)=1, \theta(B)=1-\theta_1 B$.

2. (0,2,2) 过程:

$$\begin{aligned}\nabla^2 z_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}\\ &= (1-\theta_1 B-\theta_2 B^2)z_t\end{aligned}$$

对应于 $p=0, d=2, q=2, \phi(B)=1, \theta(B)=1-\theta_1 B-\theta_2 B^2$ 。

3. (1,1,1) 过程:

$$\nabla z_t - \phi_1 \nabla z_{t-1} = a_t - \theta_1 a_{t-1}$$

或

$$(1-\phi_1 B)\nabla z_t = (1-\theta_1 B)a_t$$

对应于 $p=1, d=1, q=1, \phi(B)=1-\phi_1 B, \theta(B)=1-\theta_1 B$ 在描述非平稳时间序列时 (季节模型在第 9 章中考虑), 我们几乎很少遇到 p, d 或 q 大于 2 的情形。通常这些系数的值取 0 或 1 就合适了. 例如, 稍后我们将要说明, 利用表 4.2 所示简单模型对图 4.2 中给出的序列 A,B,C,D 作拟合, 得到很好的效果[①] .

z 的非线性变换　如果我们允许作某些变换, 那么, 模型 (4.1.9) 的有效应用范围将得以拓展。因此, 我们可以在 (4.1.9) 中用 $z_t^{(\lambda)}$ 代替 z_t, 其中 $z_t^{(\lambda)}$ 是 z_t 的某种非线性变换, 可包含一个或多个变换参数. 如何选取适当的变换要依具体情况而定, 有时

[①] 后文我们还要进一步讨论, 在没有把握的情况下, 使用非平稳模型比平稳模型更为有利, 特别是, 前面的模型中没有一个假设 z_t 有固定的均值. 但是, 我们在第 7 章中还将说明, 在某些情形得到拟合效果稍好的平稳模型是可能的.

表 4.2 用简单非平稳模型拟合图 4.1 中时间序列的一览表

序列	模型	其它模型
A	$\nabla z_t=(1-0.7B)a_t$	(0,1,1)
B	$\nabla z_t=(1+0.1B)a_t$	(0,1,1)
C	$(1-0.8B)\nabla z_t=a_t$	(1,1,0)
D	$\nabla z_t=(1-0.1B)a_t$	(0,1,1)

可根据数据来判断. 例如, 若我们对新近引入一种商品的销售感兴趣, 我们会发现商品销售量快速增长, 且就非平稳程度而言, 百分率波动比绝对波动要稳定一些. 这些事实表明应该对销售量取对数进行分析, 从下式看出

$$\nabla \log(z_t) = \log(z_t/z_{t-1}) = \log(1+\frac{\nabla z_t}{z_{t-1}}) \simeq \frac{\nabla z_t}{z_{t-1}}$$

其中 $\nabla z_t/z_{t-1}$ 是相对变化率或百分变化率, 如果相对变化不是很大, 上式近似成立. 当数据取值跨越很大范围, 特别是对于季节数据, 利用 Box 和 Cox[38] 的方法挑选变换的类型将是很有用的 (节 9.4) .

4.2 求和自回归滑动平均模型的三种显式

我们现在考虑一般模型的三种不同的显式. 其中的每一个都允许采用某种特定方式. 因此, 过程的当前值 z_t 可表示为

1. 直接利用差分方程, 用 z 的过去值和 a 的当前值、过去值表示;

2. 只用冲击的当前值和过去值 a_{t-j} 表示;

3. 用过去值 z_{t-j} 的加权和及当前冲击 a_t 表示.

在本章中我们首要关心的是非平稳模型, 其中 $\nabla^d z_t$ 是平稳过程, d 是任意整数. 不失一般性, 对于这种模型我们在陈述时可以省去 μ, 或等价地, 用 z_t 代替 $\tilde{z}_t$. 只要把 z_t 解释为相对于均值 μ 的偏差, 那么, 本章和下一章的结果都可应用于 $d=0$ 的平稳情形.

4.2.1 模型的差分形式

直接利用差分方程，我们便可以把过程的当前值 z_t 用 z 的过去值和 a 的当前值及过去值表示，即若

$$\varphi(B)=\phi(B)(1-B)^d=1-\varphi_1B-\varphi_2B^2-\cdots-\varphi_{p+d}B^{p+d}$$

$\theta_0=0$ 时的一般模型可写成

$$z_t=\varphi_1z_{t-1}+\cdots+\varphi_{p+d}z_{t-p-d}-\theta_1a_{t-1}-\cdots-\theta_qa_{t-q}+a_t \quad (4.2.1)$$

例如，考虑阶为 (1,1,1) 的模型所表示的过程

$$(1-\phi B)(1-B)z_t=(1-\theta B)a_t$$

这里，为了方便起见，我们省略了 ϕ_1 和 θ_1 的下标 1．由此，过程可以写为

$$[1-(1+\phi)B+\phi B^2]z_t=(1-\theta B)a_t$$

即

$$z_t=(1+\phi)z_{t-1}-\phi z_{t-2}+a_t-\theta a_{t-1} \quad (4.2.2)$$

对于许多应用，特别是在第 5 章中预测值的计算，差分方程 (4.2.1) 是用起来最方便的形式。

4.2.2 模型的随机冲击形式

模型用当前和过去的冲击表示　我们在节 3.1.1 中已经看到，一个线性模型可以写成以 z_t 为输出的线性滤波器：

$$\begin{aligned} z_t &= a_t+\psi_1a_{t-1}+\psi_2a_{t-2}+\cdots \\ &= a_t+\sum_{j=1}^{\infty}\psi_ja_{t-j} \quad (4.2.3) \\ &= \psi(B)a_t \end{aligned}$$

其中输出是白噪声或互不相关的冲击序列 a_t . 把 ARIMA 模型表示成 (4.2.3) 的形式, 有时是很有用的, 比如, 在第 5 章中计算预报方差时就要用到权数 ψ . 然而, 由于非平稳 ARIMA 过程关于时间不是统计均衡的, 不能假定它们可以向无穷的过去拓展, 因此形如 (4.2.3) 的无穷表示是不可能的. 但是, 相对有穷的截尾形式总是存在的, 在后文中我们将加以讨论. 现在我们来说明, ARIMA 的 ψ 权可以直接由过程的差分方程得到.

ψ 权的一般表示　如果在 (4.2.3) 的两边用广义自回归算子 $\varphi(B)$ 作用, 我们得到

$$\varphi(B)z_t = \varphi(B)\psi(B)a_t$$

然而, 由于

$$\varphi(B)z_t = \theta(B)a_t$$

可得

$$\varphi(B)\psi(B) = \theta(B) \tag{4.2.4}$$

在下面展式中

$$\begin{aligned}(1-\varphi_1 B - \cdots - \varphi_{p+d}B^{p+d})(1+\psi_1 B+\psi_2 B^2+\cdots)\\ = (1-\theta_1 B - \cdots - \theta_q B^q)\end{aligned} \tag{4.2.5}$$

令 B 的系数相等便可得到 ψ 权. 我们看到, ARIMA 的权数 ψ_j 可以由下面方程递推得出

$$\psi_j = \varphi_1\psi_{j-1} + \varphi_2\psi_{j-2} + \cdots + \varphi_{p+d}\psi_{j-p-d} - \theta_j \quad (j>0)$$

这里, $\psi_0 = 1, \psi_j = 0$ 当 $j<0$ 及 $\theta_j = 0$ 当 $j>q$. 我们注意到, 当 j 大于 $p+d-1$ 和 q 的较大者时, 权 ψ 满足由广义自回归算子定义的差分方程, 即

$$\varphi(B)\psi_j = \phi(B)(1-B)^d\psi_j = 0 \tag{4.2.6}$$

其中 B 是作用于下标 j 的算子. 因此, 当 j 足够大时, 权 ψ_j 表示为以 j 作自变量的多项式、衰减指数和衰减正弦的混合形式.

例：为了说明 (4.2.5) 的使用，我们考虑阶数为 (1,1,1) 的过程 (4.2.2)。有

$$\begin{aligned}\varphi(B) &= (1-\phi B)(1-B) \\ &= 1-(1+\phi)B+\phi B^2\end{aligned}$$

及

$$\theta(B) = 1-\theta B$$

代入 (4.2.5) 得到

$$[1-(1+\phi)B+\phi B^2](1+\psi_1 B+\psi_2 B^2+\cdots) = 1-\theta B$$

因此，ψ_j 满足递推关系 $\psi = (1+\phi)\psi_{j-1}-\phi\psi_{j-2}, j > 2$, 其中 $\psi_0 = 1, \psi_1 = (1+\phi)-\theta$。那么，由于 $\varphi(B) = (1-\phi B)(1-B) = 0$ 的根为 $G_1^{-1} = 1$ 和 $G_2^{-1} = \phi^{-1}$, 我们得到解的一般形式为

$$\psi_j = A_0 + A_1\phi^j \tag{4.2.7}$$

这里，利用初值 $\psi_0 = A_0 + A_1 = 1$ 及 $\psi = A_0 + A_1\phi = 1+\phi-\theta$ 可将常数 A_0 和 A_1 确定为

$$A_0 = \frac{1-\theta}{1-\phi}, \qquad A_1 = \frac{\theta-\phi}{1-\phi}$$

因而，我们可将模型 (4.2.2) 非正式地表式成下面的等价形式

$$z_t = \sum_{j=0}^{\infty}(A_0 + A_1\phi^j)a_{t-j} \tag{4.2.8}$$

由于 $|\phi| < 1$, 对充分大的 j 权数 ψ_j 趋于 A_0, 因此若追溯到久远的过去冲击 a_{t-j} 的权是常数 A_0。然而，表达式 (4.2.8) 并不是严格有效的，因为该式右边的无穷和在任何意义下都不收敛，因此权数 ψ_j 不是绝对可和的，这与平稳过程的情形不同。但是，模

型冲量表示的截尾形式总是有效的，我们随后就要详细讨论。尽管如此，为了表达上的方便，我们还是经常把 ARIMA 过程的无穷随机冲量表示 (即便这种表示不是严格收敛) 看成是一种简单的标记手段，用以表示下文 (4.2.14) 式的有效截尾形式，在这里两种形式的差别并不重要。

随机冲击模型的截断形式　出于技术上的需要，我们考虑与 (4.2.3) 略有差别的模型，这样做很必要，有时也是很方便的。假设我们希望将过程的当前值 z_t 用 $t-k$ 个冲击 $a_t, a_{t-1}, \cdots, a_{k+1}$ 表示，这些冲击从某一时间原点 $k(k<t)$ 输入系统。例如，时间原点 k 可以是过程首次被观测的时刻。

一般模型

$$\varphi(B)z_t = \theta(B)a_t \tag{4.2.9}$$

是一个差分方程，其解为

$$z_t = C_k(t-k) + I_k(t-k) \tag{4.2.10}$$

附录 A4.1 中给出了有关线性差分方程的简要讨论。请读者注意，这样的差分方程的解和线性微分程的解是严格平行的。余函数 $C_k(t-k)$ 就是差分方程的通解，即

$$\varphi(B)C_k(t-k) = 0 \tag{4.2.11}$$

一般，通解是由某些时间函数的线性组合构成。这些函数是幂次项 t^j，实几何指数项 G^t，复几何指数项 $D^t\sin(2\pi f_0 t + F)$，其中，常数 G, f_0, 和 F 都是模型参数 (ϕ, θ) 的函数。这些项的线性组合系数可以用时刻 $k+1$ 以前的过程值所定义的初始条件来确定。特解 $I_k(t-k)$ 是满足下式的任意函数：

$$\varphi(B)I_k(t-k) = \theta(B)a_t \tag{4.2.12}$$

特别需要注意的是，在表达式中算子 B 作用于 t 而不是 k。在附录 A4.1 中证明了，当 $t-k>q$ 时

$$I_k(t-k) = \sum_{j=0}^{t-k-1} \psi_j a_{t-j} = a_t + \psi_1 a_{t-1} + \cdots + \psi_{t-k} a_{k+1} \quad (t>k) \tag{4.2.13}$$

满足前一方程，且当 $t<k$ 时，$I_k(t-k)=0$. 因此，特解 $I_k(t-k)$ 给出了无穷冲击表示 (4.2.3) 的截尾形式，而余函数则包含了过程 z 的初始特征，从这种意义上看，$C_k(t-k)$ 在时刻 $k+1$ 就已经被确定了. 因此，对于 ARIMA 过程 (4.1.3) 而言，随机冲击模型的截断形式由下式给出

$$z_t=\sum_{j=0}^{t-k-1}\psi_j a_{t-j}+C_k(t-k) \tag{4.2.14}$$

为了说明问题，我们考查图 4.4 . 前述讨论表明，任意观测 z_t 可以看作与任何既往时刻 k 有关，且被分成相加的两个部分. 第一部分 $C_k(t-k)$ 是在时刻 k 就已确定了的 z_t 的分量，这部分表明，只要有了 $k+1$ 时刻之前的观测就能预报 t 时刻的序列值，它还表示若在 k 时刻冲击 a_t 被“切断”，过程将会采取的态势. 第二部分 $I_k(t-k)$ 表示一个在 k 时刻不可预报的附加分量，它体现了在 k 时刻之后进入系统的冲击的全部作用. 因此，为了说明一个 ARIMA 过程，必须关于某个有限过去 (当然，可能是久远的过去) 的时间原点 k 来确定 (4.2.14) 中的初始化分量 $C_k(t-k)$,

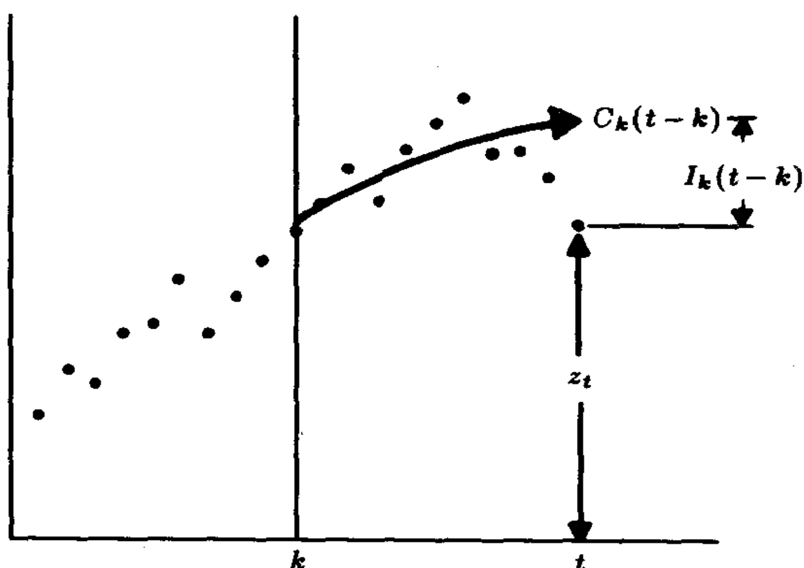


图 4.4 余函数 $C_k(t-k)$ 和特解 $I_k(t-k)$ 在描述时间序列特性中的作用

过程余下的态势由 (4.2.14) 中随机冲击的截断表示决定。

例：为了进一步说明，我们再次考虑例子

$$(1-\phi B)(1-B)z_t=(1-\theta B)a_t$$

余函数是下面差分方程的解

$$(1-\phi B)(1-B)C_k(t-k)=0$$

即

$$C_k(t-k)=b_0^{(k)}+b_1^{(k)}\phi^{t-k}$$

这里，系数 $b_0^{(k)},b_1^{(k)}$ 是依赖于过程历史值的系数，并标明它们是随原点 k 而改变的。

利用 (4.2.7) 式的 ψ 权，特解 (4.2.13) 是

$$I_k(t-k)=\sum_{j=0}^{t-k-1}(A_0+A_1\phi^j)a_{t-j}$$

最后我们可以将模型 (4.2.8) 写成下面的等价形式：

$$z_t=b_0^{(k)}+b_1^{(k)}\phi^{t-k}+\sum_{j=0}^{t-k-1}(A_0+A_j\phi^j)z_{t-j} \tag{4.2.15}$$

注意，由于 $|\phi|<1$，故当 $t-k$ 充分大时，(4.2.15) 中含有 ϕ^j 的项很小，因而可以被忽略。

随机冲击模型的截断形式与非截断形式之间的联系　现在回到一般情形。提到过程，我们可能总是联想到某个有限 (可能是久远) 的时间原点，并具有 (4.2.14) 中所示的截断随机冲击表示。通过与 (4.2.3) 的非截断形式对比，我们看到，可以把余函数 $C_k(t-k)$ 非正式地对应表示为 ψ 权的形式

$$C_k(t-k)=\sum_{j=t-k}^{\infty}\psi_j a_{t-j} \tag{4.2.16}$$

尽管从形式上看，(4.2.16) 右边的无穷和是不收敛的，但如前文所述，为了表示方便，我们还是常常使用相应的表示。

对于一般模型 (4.2.9) 总结如下：

1. 我们可以将过程 z_t 的值非正式地表示为当前和过去冲击 a_{t-k} 的无穷加权和，相应表示为

$$z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B) a_t$$

2. z_t 的值可以正式地表示为：某原点 k 之后发生的 $t-k$ 个当前和过去冲击的有穷加权和，再加上一个余函数。这个有穷和是由无穷和的前 $t-k$ 项组成的，即

$$z_k = C_k(t-k) + \sum_{j=0}^{t-k-1} \psi_j a_{t-j} \tag{4.2.17}$$

最后为了表示方便，余函数 $C_k(t-k)$ 可以采取截断无穷和的表示形式：

$$C_k(t-k) = \sum_{j=t-k}^{\infty} \psi_j a_{t-j} \tag{4.2.18}$$

为了进一步说明，我们再来考虑模型

$$(1-\phi B)(1-B) z_t = (1-\theta B) a_t$$

我们可以将 z_t 写成 a_{t-j} 无穷加权和的非正式形式：

$$z_t = \sum_{j=0}^{\infty} (A_0 + A_1 \phi^j) a_{t-j}$$

或者更正式地表示为有穷和形式：

$$z_t = C_k(t-k) + \sum_{j=0}^{t-k-1} (A_0 + A_1 \phi^j) a_{t-j}$$

此外，余函数 $C_k(t-k)$ 可以写成

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)}\phi^{t-k}$$

确定 $b_0^{(k)}, b_1^{(k)}$ 使其满足 k 时刻的初始条件，则有

$$b_0^{(k)} = \frac{z_k - \phi z_{k-1} - \theta a_k}{1-\phi}, \quad b_1^{(k)} = \frac{-\phi(z_k - z_{k-1}) + \theta a_k}{1-\phi}$$

余函数也可以非正式地表示为如下的截断无穷和

$$C_k(t-k) = \sum_{j=t-k}^{\infty} (A_0 + A_1\phi^j) a_{t-j}$$

由此能够看出 $b_0^{(k)}$ 和 $b_1^{(k)}$ 可表示为

$$b_0^{(k)} = A_0 \sum_{j=t-k}^{\infty} a_{t-j} = \frac{1-\theta}{1-\phi} \sum_{j=t-k}^{\infty} a_{t-j}$$

$$b_1^{(k)} = A_1 \sum_{j=t-k}^{\infty} \phi^{j-(t-k)} a_{t-j} = \frac{\theta-\phi}{1-\phi} \sum_{j=t-k}^{\infty} \phi^{j-(t-k)} a_{t-j}$$

余函数作为条件期望　截断表示 (4.2.14) 的一个结果是当 $m > 0$ 时有

$$C_k(t-k) = C_{k-m}(t-k+m) + \psi_{t-k}a_k + \psi_{t-k+1}a_{k-1} + \cdots + \psi_{t-k+m-1}a_{k-m} \tag{4.2.19}$$

该式表明了余函数是怎样随着原点 k 的改变而变化. 用 $E_k[z_t]$ 记 z_t 在时刻 k 的条件期望，这是在给出直到 (但不超过)k 时刻的全部历史信息条件下的期望. 为了计算条件期望值，注意

$$E_k[a_j] = \begin{cases} 0, & j > k \\ a_j, & j \leqslant k \end{cases}$$

该式说明，如果立足于时刻 k,a 将要发生的预期值为零，已发生的期望就是实际的实现值。因此，当 $(t-k)>q$ 时，余函数给出了从时刻 k 出发并在过去信息的基础上对过程未来值 z_t 的预期值。特解则表明，通过由冲击 $a_{k+1},a_{k+2},\cdots,a_t$ 所描述的后续事件，如何用来对期望进行修正。在第 5 章有关预报问题的讨论中，我们要说明， $C_k(t-k)$ 是在时刻 k 所作的 z_t 的最小方差预报。方程 (4.2.19) 可以用来对预报作修正。

4.2.3 模型的逆转形式

模型用以往 z 和当前冲击 a_t 的表示　在节 3.1.1 中我们已经看到，模型

$$z_t=\psi(B)a_t$$

可以写成下面的逆转形式：

$$\psi^{-1}(B)z_t=a_t$$

或

$$\pi(B)z_t=(1-\sum_{j=1}^{\infty}\pi B^j)z_t=a_t \tag{4.2.20}$$

因此， z_t 表为 z 的以往值再加上一个随机冲击

$$z_t=\pi_1z_{t-1}+\pi_2z_{t-2}+\cdots+a_t$$

由于可逆性条件， (4.2.20) 中的 π 权必然表示一个收敛级数，即 $\pi(B)$ 必定在单位圆内或圆上收敛。

π 权的一般表示　为了推出一般 ARIMA 模型的 π 权，我们可以将 (4.2.20) 代入下式

$$\varphi(B)z_t=\theta(B)a_t$$

得到

$$\varphi(B)z_t=\theta(B)\pi(B)z_t$$

为了得出 π 权，只须在下式中

$$\varphi(B)=\theta(B)\pi(B) \tag{4.2.21}$$

令 B 的系数相等。即

$$\begin{aligned}&(1-\varphi_1 B-\cdots-\varphi_{p+d}B^{p+d})\\&\quad=(1-\theta_1 B-\cdots-\theta_q B^q)(1-\pi_1 B-\pi_2 B^2-\cdots)\end{aligned} \tag{4.2.22}$$

我们看到，过程 ARIMA 的权数 π_j 可由下式递推地确定：

$$\pi_j=\theta_1\pi_{j-1}+\theta_2\pi_{j-2}+\cdots+\theta_q\pi_{j-q}+\varphi_j \qquad (j>0)$$

其中，约定 $\pi_0=-1,\pi_j=0$ 对 $j<0$, 及 $\varphi_j=0$, 对 $j>p+d$。我们注意到，当 j 大于 $p+d$ 和 q 中的较大者时，π 权满足滑动平均算子所定义的差分方程：

$$\theta(B)\psi_j=0$$

因此当 j 足够大时，π 权表现出与自回归过程的自相关函数 (3.2.5) 类似的特性，即遵从混合的指数衰减及正弦振荡衰减。

另一个有趣的事实是：当 $d>1$ 时，(4.2.20) 式中的 π 权之和为 1, 通过在 (4.2.21) 中将 $B=1$ 代入可以得到验证。由于 $B=1$ 时 $\varphi(B)=\phi(B)(1-B)^d$ 为零，且因 $\theta(B)=0$ 的根在单位圆外，有 $\theta(1)\neq 0$, 故由 (4.2.21) 知 $\pi(1)=0$, 即

$$\sum_{j=1}^{\infty}\pi_j=1 \tag{4.2.23}$$

因此，当 $d>1$ 时，过程可以写成如下形式：

$$z_t=\bar{z}_{t-1}(\pi)+a_t \tag{4.2.24}$$

其中

$$\bar{z}_{t-1}(\pi)=\sum_{j=1}^{\infty}\pi_j z_{t-j}$$

是过程以往值的加权平均.

例： 为了说明起见，我们再次考虑 ARIMA(1,1,1) 过程

$$(1-\phi B)(1-B)z_t=(1-\theta B)a_t$$

那么，利用 (4.2.21), 有

$$\pi(B)=\varphi(B)\theta^{-1}(B)=[1-(1+\phi)B+\phi B^2](1+\theta B+\theta^2B^2+\cdots)$$

因此

$$\pi_1=\phi+(1-\theta),\qquad \pi_2=(\theta-\phi)(1-\theta)$$
$$\pi_j=(\theta-\phi)(1-\theta)\theta^{j-2}\quad (j>3)$$

作为例子，表 4.3 给出了 $\phi=-0.3,\theta=0.5$ 的 ARIMA 模型 (1,1,1) 的前 7 个 π 权的值. 因此，过程 t 时刻的值可由以往值的加权平均再加上一个冲击生成，即相应于

$$z_t=(0.2z_{t-1}+0.4z_{t-2}+0.2z_{t-3}+0.1z_{t-4}+\cdots)+a_t$$

我们注意到，随着相应的 z_{t-j} 越远，π 权越衰减. 这个性质是由于要求序列为可逆而引起的 (这里要求 $-1<\theta<1$) .

先前我们曾提到，当用统计模型描述实际中出现的时间序列时，收敛的 π 权通常衰减得很快. 因此，尽管从理论上 z_t 可能依赖于久远的过去，但是，表达式

$$z_t=\sum_{j=1}^{\infty}\pi_j z_{t-j}+a_t$$

表明，z_t 往往只依赖于时间序列近期过去这一重要范围内的 z_{t-j} . 对于 $d>0$ 的非平稳模型也是如此. 加权“冲击”表示式

$$z_t=\sum_{j=0}^{\infty}\psi_j a_{t-j}$$

表 4.3 $\phi=-0.3,\theta=0.5$ 的 ARIMA(1,1,1) 过程对应 π 权中的前 7 个值

j	1	2	3	4	5	6	7
π	0.2	0.4	0.2	0.1	0.05	0.025	0.0125

其中 ψ 权并不收敛。当然也存在这种可能，即冲击 a_{t-j} 的久远过去值提供给 z_t 的全部信息都包含在序列的近期值 $z_{t-1}, z_{t-2}, \cdots$ 中。特别，条件期望 $E_k[z_t]$ 是以直到 k 时刻为止的全部历史信息为条件的，从理论上它可以用时间序列的近期值计算得足够精确，这一事实在预报应用中特别重要。

4.3 求和滑动平均过程

非平稳过程 (0,1,1) 在表示经常出现的序列时是有用的，即是

$$\nabla z_t = a_t - \theta a_{t-1}$$

该模型只含两个参数：θ 和 σ_a^2。图 4.5 中给出了两个时间序列，它们是由同一正态偏差序列 a_t 按上述模型生成的，第一个序列的 $\theta=0.6$，第二个序列 $\theta=0$。人们发现有关库存控制问题，生产过程中出现的某类干扰以及在计量经济等方面，这种模型都是很有用的。正如节 4.1.3 所述，在第 7 章中将要说明：选取适当

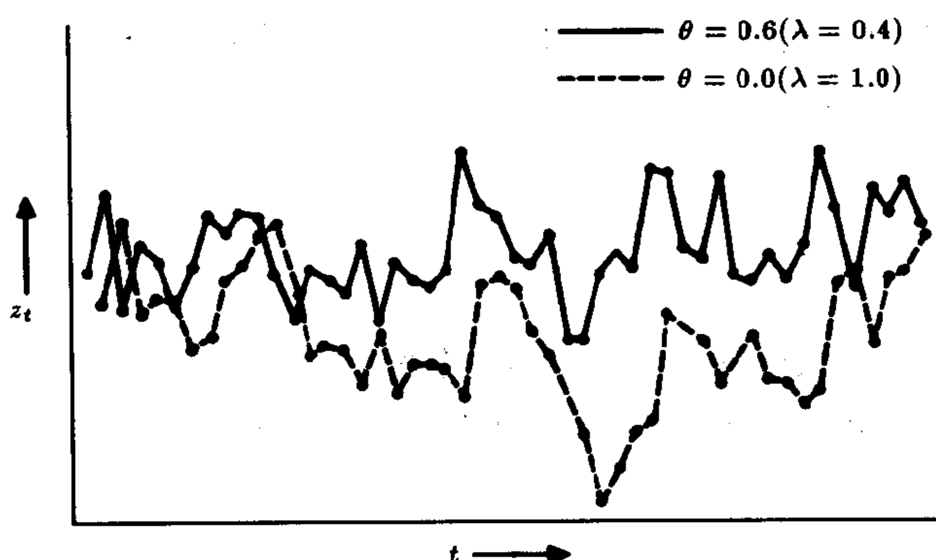


图 4.5 模型 IMA(0,1,1) 生成的两个时间序列

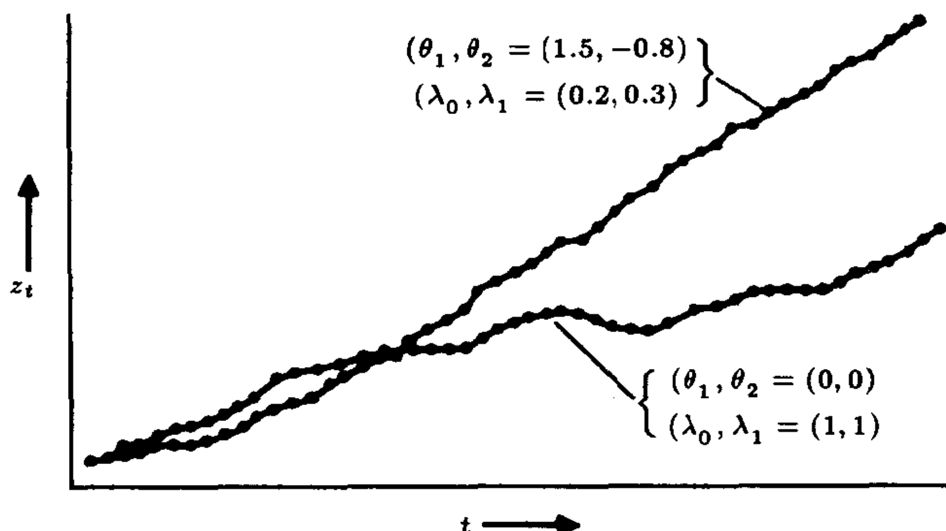


图 4.6 IMA(0,2,2) 模型生成的两个时间序列

参数，利用这种简单的模型就可以给出图 4.1 中 A,B 和 C 序列的有用的表示．另一种重要的模型是

$$\nabla^2 z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

其中包含三个参数： θ_1, θ_2 和 σ_a ．图 4.6 给出了两个时间序列，它们也是由前文中的同一正态偏差序列由该模型生成。其中第一个模型的参数 $(\theta_1, \theta_2) = (0, 0)$，第二个模型的参数是 $(\theta_1, \theta_2) = (1.5, -0.8)$ 。这些序列比由过程 (0,1,1) 生成的序列要平滑一些。(0,2,2) 模型用于表示具有惯性较大系统中的干扰是有用的 (如序列 C) 。模型 (0,1,1) 和 (0,2,2) 都是模型

$$\nabla^d z_t = \theta(B) a_t \tag{4.3.1}$$

的特例。我们称形如 (4.3.1) 的这类模型为求和滑动平均 (IMA) 过程。

4.3.1 (0,1,1) 阶求和滑动平均过程

差分方程形式 (0,1,1) 过程

$$\nabla z_t = (1 - \theta B) z_t \qquad (-1 < \theta < 1)$$

具有很好的表现能力，现在我们来更具体地研究其特点。该模型可以写成用 z 和 a 来表示

$$z_t = z_{t-1} + a_t - \theta a_{t-1} \tag{4.3.2}$$

模型的随机冲击形式　另一方面，通过对 (4.3.2) 两边求和，我们可以得到 z_t 只用 a 表示的形式。为此，在上式右边用算子 ∇ 而不用 B 会更方便一些，因此我们写为

$$1 - \theta B = (1-\theta)B + (1-B) = (1-\theta)B + \nabla = \lambda B + \nabla$$

这里， $\lambda = 1 - \theta$, 用 λ 定义的可逆域定义为 $0 < \lambda < 2$. 于是

$$\nabla z_t = \lambda a_{t-1} + \nabla a_t$$

相对于某一时间原点 $k < t$, 利用有限和算子 $S_{t-k} = 1 + B + \cdots + B^{t-k-1} = (1-B^{t-k})/(1-B)$, 我们得到

$$(1-B^{t-k})z_t = \lambda S_{t-k} a_{t-1} + (1-B^{t-k})a_t \tag{4.3.3}$$

故

$$z_t = a_t + \lambda(a_{t-1} + a_{t-2} + \cdots + a_{k+1}) + (z_k - \theta a_k) \tag{4.3.4}$$

对比 $z_t = \sum_{j=0}^{t-k-1} \psi_j a_{t-j} + C_k(t-k)$, 权数为 $\psi_0 = 1, \psi_j = \lambda$, 对 $j \geqslant 1$. 另外，余函数 $C_k(t-k) = z_k - \theta a_k = b_0^{(k)}$(对任何 k 都是常数 b_0) 是差分方程 $(1-B)C_k(t-k) = 0$ 的解。进而，在无穷表示 $z_t = a_t + \lambda \sum_{j=1}^{\infty} a_{t-j}$ 中，我们令 $b_0^{(k)}$ 与 $\lambda \sum_{j=t-k}^{\infty} a_{t-j}$ 等同。因此对于该模型，余函数简化为一个常数 (也即为 t 的零阶多项式), 以此来表示过程的当前水平及具体参考原点 k 之间的联系。如果原点从 $k-1$ 变化到 k, 则 b_0 依照下式修正

$$b_0^{(k)} = b_0^{(k-1)} + \lambda a_k$$

这里利用 (4.3.2) 式，有 $b_0^{(k)} = z_k+(\lambda-1)a_k = z_{k-1}-\theta a_{k-1}+\lambda a_k$.

模型的逆转形式　最后，我们还可以考虑模型的如下形式：

$$\pi(B)z_t = a_t$$

或等价地表示为

$$z_t = \sum_{j=1}^{\infty}\pi_j z_{t-j} + a_t = \bar{z}_{t-1}(\pi) + z_t$$

这里，$\bar{z}_{t-1}(\pi)$ 是过程既往值的加权滑动平均。

利用 (4.2.21), 过程 IMA(0,1,1) 的 π 权由下式给出

$$(1-\theta B)\pi(B) = 1 - B$$

即

$$\begin{aligned}\pi(B) &= \frac{1-B}{1-\theta B} = \frac{1-\theta B-(1-\theta)B}{1-\theta B} \\ &= 1-(1-\theta B)(B+\theta B^2+\theta^2 B^3+\cdots)\end{aligned}$$

因此

$$\pi_j = (1-\theta)\theta^{j-1} = \lambda(1-\lambda)^{j-1} \qquad j > 1$$

于是过程可以写成

$$z_t = \bar{z}_{t-1}(\lambda) + a_t \tag{4.3.5}$$

过程以往值的滑动加权平均

$$\bar{z}_{t-1}(\lambda) = \lambda\sum_{j=1}^{\infty}(1-\lambda)^{j-1}z_{t-j} \tag{4.3.6}$$

在这里是指数形式的加权滑动平均 (EWMA) .

(4.3.6) 的滑动平均称为指数 (或几何) 加权，这是由于权数

$$\lambda\,, \qquad \lambda(1-\lambda)\,, \qquad \lambda(1-\lambda)^2\,, \qquad \lambda(1-\lambda)^3 \;\cdots$$

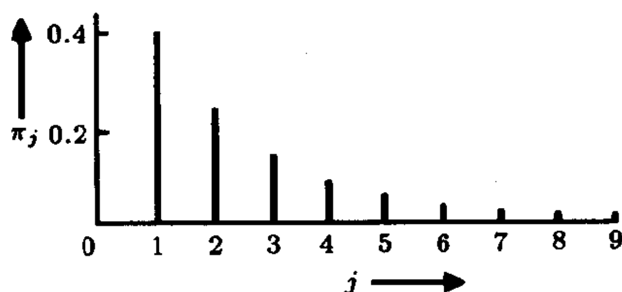


图 4.7 参数 $\lambda=1-\theta=0.4$ 的 (0,1,1) 阶 IMA 过程

呈指数 (即作为一个几何级数) 衰减. 在图 4.7 中给出了一个阶数为 (0,1,1) 的 IMA 过程的权函数, 其中 $\lambda=0.4(\theta=0.6)$ 。

尽管可逆性条件是满足 $0<\lambda<2$, 但是在实际中我们经常考虑的是 λ 在 0 和 1 之间的情形. 我们注意到, 如果 λ 的值为 1, 则权函数由单一的峰值构成 ($\pi_1=1, \pi_j=0$, 对 $j>1$) . 当 λ 的值接近于 0 时, 指数权收敛得越来越慢, $EWMA$ 使过去的值伸延到未来. 最后, 当 $\lambda=0$ 时, $\theta=1$, 模型 $(1-B)z_t=(1-B)a_t$ 等价于 $z_t=\theta_0+a_t$, 其中 θ_0 由过去值的均值给出.

由于 $b_0^{(k)}=z_k-\theta a_k=z_{k+1}-a_{k+1}$, 或 $z_{k+1}=b_0^{(k)}+a_{k+1}$, 对比 (4.3.5) 可得, 对于该过程, (4.3.5) 中的余函数 $b_0^{(k)}=C_k(t-k)$ 为

$$b_0^{(k)}=\bar{z}_k(\lambda) \tag{4.3.7}$$

这是直到原点 k 的所有值的指数加权平均. 事实上, (4.3.4) 可以写成

$$z_t=\bar{z}_k(\lambda)+\lambda\sum_{j=1}^{t-k-1}a_{t-j}+a_t$$

我们曾经说过,若立足于时刻 k 并基于过去信息,那么,余函数可以看作是向我们提供过程在 t 时刻的未来值. 对于 IMA(0,1,1) 过程来说, 所采取的信息形式是关于过程的水平或位置 $b_0^{(k)}=\bar{z}_k(\lambda)$. 在时刻 k 我们具有关于过程未来特性的信息, 但这将会导致以 "随机游动" 方式从水平偏离, 我们表示为 $\lambda\sum_{j=1}^{t-k-1}a_{t-j}+a_t$, 它

的期望为零且其有关特性是不可预报的。一旦得到新的观测值，我们立刻把原点移到时刻 $k+1$，同时水平修正为 $b_0^{k+1}=\bar{z}_{k+1}(\lambda)$ 。

IMA(0,1,1) 过程的重要性质　由于非平稳性，因此过程并不是以一种稳定方式围绕固定的均值变动。但是指数加权滑动平均 $\bar{z}_t(\lambda)$ 可以看作是过程在时刻 t 的位置或"水平"的度量。从其定义 (4.3.6) 式我们得到著名的 EWMA 递推公式

$$\bar{z}_t(\lambda)=\lambda z_t+(1-\lambda)\bar{z}_{t-1}(\lambda) \tag{4.3.8}$$

该式表明，对于 $EMA(0,1,1)$ 过程来说，每一个新水平都是通过对新观测和先前水平的插值得到。如果 λ 等于 1，那么 $\bar{z}_t(\lambda)=z_t$，这表明可以忽略与过去观测值有关的所有信息。另一方面，如果 λ 的值接近于 0，过去值的权为 $1-\lambda$，$\bar{z}_t(\lambda)$ 将主要于依赖过去值 $\bar{z}_{t-1}(\lambda)$，只有很小的权数给予新观测值。

现在考虑下面的两个方程

$$\begin{aligned} z_t &= \bar{z}_{t-1}(\lambda)+a_t \\ \bar{z}_t(\lambda) &= \bar{z}_{t-1}(\lambda)+\lambda a_t \end{aligned} \tag{4.3.9}$$

后一方程是通过把 (4.3.5) 代入 (4.3.8) 得到的，也可以由 (4.3.7) 直接推出。

Muth[147] 曾指出，(4.3.9) 的两个方程给出了考虑过程演化的有效手段。第一个方程表明若系统的水平为 $\bar{z}_{t-1}(\lambda)$，由于在 t 时刻加入了冲击 a_t，因而生成了观测值 z_t 。第二个方程表明，只有比例为 λ 的冲击真正被水平所吸收并具有持续的影响，冲击的其余部份 $\theta=1-\lambda$ 消失了。当 $t+1$ 时刻新的冲击输入时，系统已经通过吸收 a_t 确立了新的水平。将方程 (4.3.9) 的下标增加 1，它便用来描述冲击是如何产生 z_{t+1}，及冲击的一部分 (λ) 是如何被系统吸收并产生新的水平 $\bar{z}_{t+1}(\lambda)$，如此等等。

利用方程 (4.3.4) 可以直接得到 IMA(0,1,1) 的方差和相关特性。例如，在以 k 为原点时，初始函数 $b_0^{(k)}$ 可以看作是常数，我

们求得

$$\text{var}[z_t] = \sigma_a^2[1 + (t - k - 1)\lambda^2] \tag{4.3.10}$$

当 t 增大时上式不收敛. 事实上, 我们可以把这一方差看作是差分 $z_t - z_k$ 的方差, (在 (4.3.4) 中取 $a_k = 0$). 特别, 当过程为随机游动 $z_t = z_{t-1} + a_t$ 的情形, 这时有 $\lambda = 1$ 且方差函数按比例 $t - k$ 增长. 反之, 对于 $0 < \lambda < 1$ 的一般情形, 特别是 λ 接近于 0 时, $z_t - z_k$ 的方差函数随 $t - k$ 增长得很慢. 另外, 对 $s > 0$ 有 $\text{cov}[z_t, z_{t+s}] = \sigma_a^2[\lambda + (t - k - 1)\lambda^2]$, 这意味着当 $t - k$ 相对于 s 很大且 λ 不接近于零时, $\text{corr}[z_t, z_{t+s}]$ 将接近于 1. 因此, 过程的相邻值之间表现出很强的正相关, 且过程呈现较为平滑的明显特性 (除非 λ 接近于零).

具有确定性偏差的 IMA(0,1,1) 过程为

$$\nabla z_t = \theta_0 + (1 - \theta_1 B)a_t$$

其性质将在附录 A4.2 中讨论.

4.3.2 (0,2,2) 阶求和滑动平均模型

差分方程形式　过程

$$\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2)a_t \tag{4.3.11}$$

使得我们能够对具有随机趋势的序列加以描述 (例如, 见图 4.6), 现在我们在可逆域

$$-1 < \theta_2 < 1 \qquad \theta_2 + \theta_1 < 1 \qquad \theta_2 - \theta_1 < 1$$

中研究该过程的一般性质. 如前所述, 用 z 和 a 可以将 z_t 写为如下显式:

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

另一方面, 在我们得到 z_t 关于 a 的表示之前, 先将上式右边的运算改写成差分形式:

$$1 - \theta_1 B - \theta_2 B^2 = (\lambda_0 \nabla + \lambda_1)B + \nabla^2$$

令系数相等，我们得到 θ 关于 λ 的表达式，同样也有相反的表达式，如下

$$\begin{aligned} \theta_1 &= 2-\lambda_0-\lambda_1 , & \lambda_0 &= 1+\theta_2 \\ \theta_2 &= \lambda_0 - 1 , & \lambda_1 &= 1-\theta_1-\theta_2 \end{aligned} \tag{4.3.12}$$

模型 (4.3.11) 可改写为

$$\nabla^2 z_t = (\lambda_0 \nabla + \lambda_1) a_{t-1} + \nabla^2 a_t \tag{4.3.13}$$

与 (4.3.11) 相比，采用模型的 (4.3.13) 形式具有很大的优越性，这是基于下面的事实：若我们在 (4.3.13) 中令 $\lambda_1 = 0$, 则可得

$$\nabla z_t = [1-(1-\lambda_0)B]a_t$$

这对应于 $\theta = 1-\lambda_0$ 的 (0,1,1) 过程．然而，若在 (4.3.11) 中令 $\theta_2 = 0$, 我们得到

$$\nabla^2 z_t = (1-\theta_1 B)a_t$$

稍后在第 5 章中我们将说明，对一个由 (0,2,2) 过程生成的序列而言，其最优预报总是沿着一条直线，随着新数据的获取，直线的水平和斜率不断修正．与此对比，由 (0,1,1) 过程生成的序列不能提供关于斜率的任何信息，仅能提供不断修正的水平值．不论对线性趋势还是水平都能作预报和修正，这是一个很重要的问题．如果要在这两个模型中挑选的话，前述问题取决于 (4.3.13) 中 λ 是否为零．

IMA(0,2,2) 过程的可逆域和在第 3 章中给出的 MA(2) 的可逆域是一样的．用 θ 或 λ 可将其写为如下形式

$$\begin{aligned} \theta_2+\theta_1 &< 1 , & 0 < 2\lambda_0 + \lambda_1 &< 4 \\ \theta_2-\theta_1 &< 1 , & \lambda_1 &> 0 \\ -1<\theta_2 &< 1 , & \lambda_0 &> 0 \end{aligned} \tag{4.3.14}$$

θ 的三角形的可逆域在图 3.8 中给出，而 λ 的相应区域如图 4.8 所示．

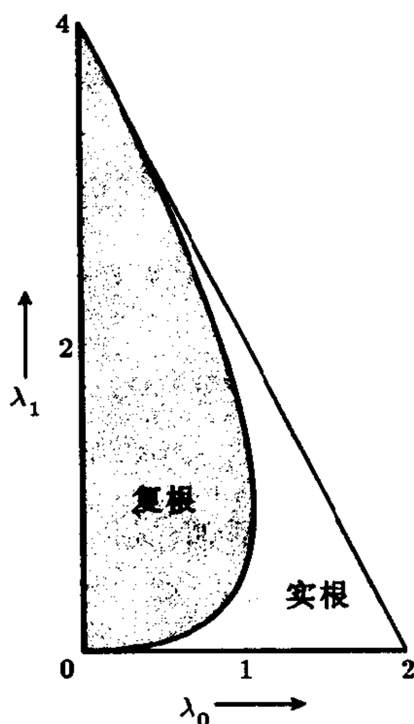


图 4.8 IMA(0,2,2) 过程关于参数 λ_0 和 λ_1 的可逆域

模型的截断及随机冲击形式　相对于时间原点 k, 对 (4.3.14) 运用有限双重和算子 $S_{t-k}^{(2)}$ 我们得出

$$
\begin{aligned}
&[1-B^{t-k}-(t-k)B^{t-k}(1-B)]z_t \\
&\quad=[\lambda_0(S_{t-k}-(t-k)B^{t-k})+\lambda_1 S_{t-k}^{(2)}]a_{t-1} \\
&\qquad+[1-B^{t-k}-(t-k)B^{t-k}(1-B)]a_t
\end{aligned}
$$

因此，我们得到随机冲击模型的截断形式

$$
\begin{aligned}
z_t&=\lambda_0 S_{t-k-1}a_{t-1}+\lambda_1 S_{t-k}^{(2)}a_{t-1}+a_t+b_0^{(k)}+b_1^{(k)}(t-k) \\
&=\lambda_0\sum_{j=1}^{t-k-1}a_{t-j}+\lambda_1\sum_{j-1}^{t-k-1}ja_{t-j}+a_t+C_k(t-k)
\end{aligned}
\tag{4.3.15}
$$

故对于该过程，ψ 权为

$$
\psi_0=1,\quad \psi_1=(\lambda_0+\lambda_1),\ \cdots,\ \psi_j=(\lambda_0+j\lambda_1),\cdots
$$

余函数是下面方程

$$
(1-B)^2C_k(t-k)=0
$$

的解，即

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)}(t-k) \tag{4.3.16}$$

这是一个 $(t-k)$ 的一阶多项式，其系数依赖于原点 k 的位置. 由 (4.3.15) 我们求出这些系数为

$$\begin{aligned} b_0^{(k)} &= z_k - (1-\lambda_0)a_k \\ b_1^{(k)} &= z_k - z_{k-1} - (1-\lambda_1)a_k + (1-\lambda_0)a_{k-1} \end{aligned}$$

另外，通过考虑差分 $b_0^{(k)} - b_0^{k(-1)}$ 和 $b_1^{(k)} - b_1^{k-1}$ 得知，若原点从 $k-1$ 改变到 k, 那么 b_0 和 b_1 按照下式

$$\begin{aligned} b_0^{(k)} &= b_0^{k-1} + b_1^{(k-1)} + \lambda_0 a_k \\ b_1^{(k)} &= b_1^{k-1} + \lambda_1 a_k \end{aligned} \tag{4.3.17}$$

予以修正. 我们看到, 如果该模型是合适的, 那么从原点 k 出发，对序列未来特性的预期可以用直线 (4.3.16) 来表示，相应的位置是 $b_0^{(k)}$, 斜率是 $b_1^{(k)}$. 特别，到时刻 t 时，过程将从该直线偏离，这是由于随机分量

$$\lambda_0 \sum_{j=1}^{t-k-1} a_{t-j} + \lambda_1 \sum_{j=1}^{t-k-1} j a_{t-j} + a_t$$

的影响，这种偏离在时刻 k 是无法预报的。此外，当时间原点从 $k-1$ 移到 k 时，相应的截距和斜率也按照 (4.3.17) 修正.

从形式上，我们也可以得到过程的无穷随机冲击表示:

$$z_t = \lambda_0 \sum_{j=1}^{\infty} a_{t-j} + \lambda_1 \sum_{j=1}^{\infty} j a_{t-j} + a_t = \lambda_0 S a_{t-1} + \lambda_1 S^2 a_{t-1} + a_t \tag{4.3.18}$$

与 (4.3.15) 式对比，余函数可以形式上地表为

$$C_k(t-k) = \lambda_0 \sum_{j=t-k}^{\infty} a_{t-j} + \lambda_1 \sum_{j=t-k}^{\infty} j a_{t-j} = b_0^{(k)} + b_1^{(k)}(t-k)$$

通过把上面第二和式表示成 $\sum_{j=t-k}^{\infty} ja_{t-j}=(t-k)\sum_{j=t-k}^{\infty} a_{t-j}+\sum_{j=t-k}^{\infty}[j-(t-k)]a_{t-j}$，我们看到，$b_0^{(k)}$ 和 $b_1^{(k)}$ 由下式给出

$$
\begin{aligned}
b_0^{(k)} &= \lambda_0 S a_k + \lambda_1 S^2 a_{k-1} = (\lambda_0 - \lambda_1) S a_k + \lambda_1 S^2 a_k \\
b_1^{(k)} &= \lambda_1 S a_k
\end{aligned}
$$

模型的逆转形式　最后，我们考虑模型的逆转形式表示

$$
z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t = \bar{z}_{t-1}(\pi) + a_t
$$

利用 (4.2.22) 式，并在下式中令两边系数相等

$$
1 - 2B + B^2 = (1 - \theta_1 B - \theta_2 B^2)(1 - \pi_1 B - \pi_2 B^2 - \cdots)
$$

得到 IMA(0,2,2) 过程的 π 权是

$$
\begin{aligned}
&\pi_1 = 2 - \theta_1 = \lambda_0 + \lambda_1 \\
&\pi_2 = \theta_1(2 - \theta_1) - (1 + \theta_2) = \lambda_0 + 2\lambda_1 - (\lambda_0 + \lambda_2)^2 \qquad (4.3.19) \\
&(1 - \theta_1 B - \theta_2 B^2)\pi_j = 0 \qquad (j > 3)
\end{aligned}
$$

这里 B 作用在 j 上.

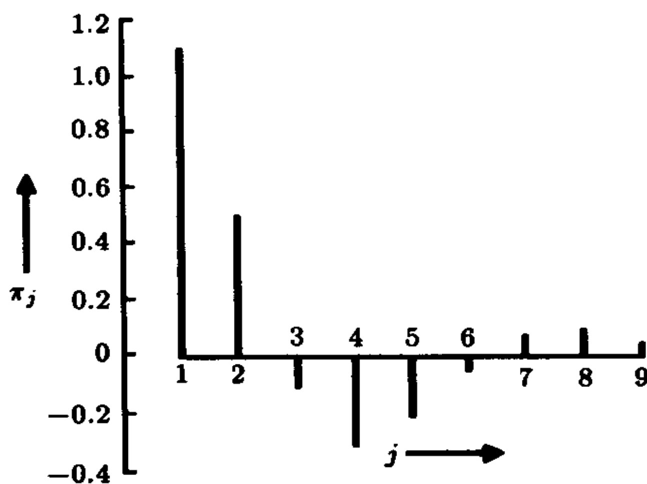


图 4.9　$\lambda_0=0.5, \lambda_1=0.6$ 的 (0,2,2) 阶 IMA 过程的 π 权

如果特征方程 $1-\theta_1 B-\theta_2 B^2$ 的根是实根，那么，π 权对过去值 z 的使用是采取两个衰减指数的混合形式，如果为复根，则 π 权遵从衰减的正弦波。图 4.9 给出了一个 IMA(0,2,2) 的权，参数为 $\theta_1=0.9, \theta=-0.5$(即 $\lambda_0=0.5, \lambda_1=0.6$)。我们从图 3.9 和图 4.8 看到，相应的特征方程有实根 (判别式 $\theta_1^2+4\theta_2^2=-1.19$ 小于 0)。因此，π 权应该遵从衰减的正弦波动，在图 3.9 中也确是如此。

4.3.3 一般的 $(0,d,q)$ 阶求和滑动平均过程

差分方程形式　一般的 $(0,d,q)$ 阶求和滑动平均过程是

$$\nabla^d z_t=(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)a_t=\theta(B)a_t \tag{4.3.20}$$

为使过程为可逆，$\theta(B)$ 的零点都必须在单位圆外。模型 (4.3.20) 可以用过去的 z 和 a 表示为如下形式

$$\begin{aligned} z_t = dz_{t-1} - \frac{1}{2}d(d-1)z_{t-2} + \cdots + (-1)^{d+1}z_{t-d} \\ + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \end{aligned}$$

模型的随机冲击形式　为了得出 z 用 a 表示，我们把 (4.3.20) 右边的运算写成 $\nabla=1-B$ 的形式，由此得到

$$(1-\theta_1 B-\cdots\theta_q B^q)=(\lambda_{d-q}\nabla^{q-1}+\cdots+\lambda_0\nabla^{d-1}+\cdots+\lambda_{d-1})B+\nabla^d \tag{4.3.21}$$

与前文相同，使上式中 B 的系数相等，我们可以将 λ 用 θ 表示。

把 (4.3.21) 代入 (4.3.20) 并进行 d 重求和，我们形式上得出

$$z_t=(\lambda_{d-q}\nabla^{q-d-1}+\cdots+\lambda_0 S+\cdots+\lambda_{d-1}S^d)a_{t-1}+a_t \tag{4.3.22}$$

当 $q>d$ 时从 (4.3.22) 中我们看到，除了 d 个和式以外，还写出了涉及 $a_{t-1}, a_{t-2}, \cdots, a_{t+d-q}$ 的 $q-d$ 个加项 $\nabla^{q-d-1}a_{t-1}, \cdots$。

如果将解表示成在某个原点时刻 k 之后输入到系统中的 a 的有穷和形式，我们得到同样的方程，只是多了一个余函数，它满足

$$\nabla^d C_k(t-k)=0$$

即为多项式

$$C_k(t-k)=b_0^{(k)}+b_j^{(k)}(t-k)+b_2^{(k)}(t-k)^2+\cdots+b_{d-1}^{(k)}(t-k)^{d-1}$$

如前所述，余函数描述了依赖于时刻 k 的过程的有限特性. 完全类似，系数 $b_j^{(k)}$ 也可以形式上地表示为直到 k 时刻的无穷和，即用 $Sa_k,S^2a_k,\cdots,S^da_k$ 表示. 相应地，当原点从 $k-1$ 变为 k 时，我们也可以看出系数 $b_j^{(k)}$ 是怎样改变的.

模型的逆转形式　最后，模型可以用逆转形式表示：

$$\pi(B)z_t=a_t$$

或

$$z_t=\bar{z}_{t-1}(\pi)+a_t$$

(4.2.22) 式成为

$$(1-B)^d=(1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q)(1-\pi_1B-\pi_2B^2-\cdots) \tag{4.3.23}$$

令系数相等即可得到 π 权. 对于一个给定的模型，得出 π 权的最好的方法是将具体数值代入 (4.3.23), 而不是去推导一般的公式. 我们注意到，当 j 大于 d 和 q 中较大者时，(4.3.23) 蕴含 π 权满足由滑动平均算子所定义的差分方程

$$\theta(B)\pi_j=0$$

因此，对于足够大的 j,π 权遵从混合的指数衰减和正弦波.

(0,2,3) 阶 IMA 过程　最后一种令人感兴趣且有价值的特殊情形是 (0,2,3) 阶 IMA 过程

$$\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3) a_t$$

如前所述，运用有限重求和运算，我们可将模型写为截断的随机冲击形式

$$z_t = \lambda_{-1} a_{t-1} + \lambda_0 \sum_{j=1}^{t-k-1} a_{t-j} + \lambda_1 \sum_{j=1}^{t-k-1} j a_{t-j} + a_t + b_0^{(k)} + b_1^{(k)}(t-k)$$

λ 和 θ 之间的对应关系如下：

$$\begin{aligned} \theta_1 &= 2 - \lambda_{-1} - \lambda_0 - \lambda_1 , & \lambda_{-1} &= -\theta_3 \\ \theta_2 &= \lambda_0 - 1 + 2\lambda_{-1} , & \lambda_0 &= 1 + \theta_2 + 2\theta_3 \\ \theta_3 &= -\lambda_{-1} , & \lambda_1 &= 1 - \theta_1 - \theta_2 - \theta_3 \end{aligned}$$

另外，也可以形式上些写成下面的无限求和形式：

$$z_t = \lambda_{-1} a_{t-1} + \lambda_0 S a_{t-1} + \lambda_1 S^2 a_{t-1} + a_t$$

最后，可逆域定义为

$$\begin{aligned} \theta_1 + \theta_2 + \theta_3 &< 1 , & \lambda_1 &> 0 \\ -\theta_1 + \theta_2 - \theta_3 &< 1 , & 2\lambda_0 + \lambda_1 &< 4(1 - \lambda_{-1}) \\ \theta_3(\theta_3 - \theta_1) - \theta_2 &< 1 , & \lambda_0(1 + \lambda_{-1}) &> -\lambda_1 \lambda_{-1} \\ |\theta_3| &< 1 , & |\lambda_{-1}| &< 1 \end{aligned}$$

在第 5 章中我们将要说明，若模型为 ARIMA 过程，如何采用一种最优的方法对时间序列的未来值进行预报。

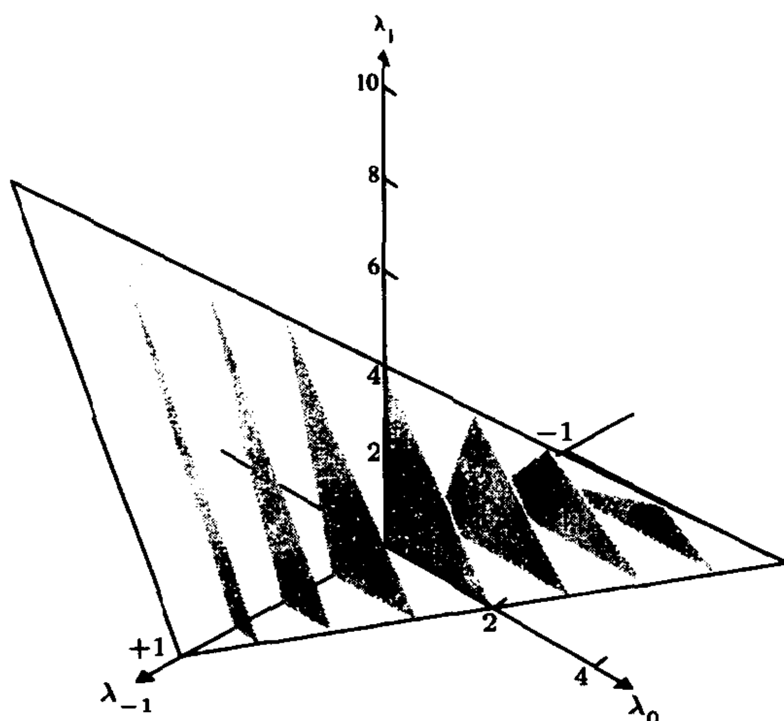


图 4.10 参数为 λ_{-1}, λ_0 及 λ_1 的 IMA(0,2,3) 模型可逆域

A4.1 线性差分方程

在本书中，我们经常涉及线性差分方程。特别， ARIMA 模型是用如下形式的差分方程

$$\begin{aligned} z_t - \varphi_1 z_{t-1} - \varphi_2 z_{t-2} - \cdots - \varphi_{p'} z_{t-p'} \\ = z_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \end{aligned} \tag{A4.1.1}$$

这里 $p' = p + d$ 。

另一方面，我们可将 (A4.1.1) 写为

$$\varphi(B) z_t = \theta(B) a_t$$

这里

$$\begin{aligned} \varphi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p'} B^{p'} \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \end{aligned}$$

我们来推导差分方程 (A4.1.1) 关于时间原点 $k < t$ 的通解表达式。

1. 我们证明通解可以写为

$$z_t = C_k(t-k) + I_k(t-k)$$

这里， $z_t = C_k(t-k)$ 是余函数， $I_k(t-k)$ 是特解。

2. 然后我们推导余函数 $C_k(t-k)$ 的一般表达式。

3. 最后，我们推出特解 $I_k(t-k)$ 的一般表达式。

通解 对于线性差分方程或线性代数方程的解进行讨论都是同样的。假设 z'_t 是

$$\varphi(B)z_t = \theta(B)a_t \tag{A4.1.2}$$

的特解，即它满足

$$\varphi(B)z'_t = \theta(B)a_t \tag{A4.1.3}$$

因而 $z''_t = z_t - z'_t$ 满足

$$\varphi(B)z''_t = 0 \tag{A4.1.4}$$

于是有

$$z_t = z''_t + z'_t$$

因此 (A4.1.2) 的通解是余函数 z''_t(齐次差分方程 (A4.1.4) 的通解) 加上特解 z'_t((A4.1.2) 的任一特解) 。相对于任一时间原点 $k < t$, 我们把余函数 z''_t 记为 $C_k(t-k)$, 把特解 z'_t 记为 $I_k(t-k)$ 。

余函数的计算

无重根 考虑齐次差分方程

$$\varphi(B)z_t = 0 \tag{A4.1.5}$$

其中

$$\varphi(B) = (1-G_1B)(1-G_2B)\cdots(1-G_{p'}B) \tag{A4.1.6}$$

首先我们假定 $G_1, G_2, \cdots, G_{p'}$ 是互不相同的。下文将要证明，若序列以时刻 k 为原点，则 (A4.1.5) 在时刻 t 的通解是

$$z_t = A_1G_1^{t-k} + A_2G_2^{t-k} + \cdots + A_{p'}G_{p'}^{t-k} \tag{A4.1.7}$$

其中 A_i 为常数。因此，$\varphi(B)=0$ 的一个实根就为余函数提供一个衰减指数项 G^{t-k}，一对复根就提供一个衰减正弦项 $D^{t-k}\sin(2\pi f_0 t+F)$。

为了验证 (A4.1.7) 满足 (A4.1.5)，我们将 (A4.1.7) 代入 (A4.1.5) 得出

$$\varphi(B)(A_1G_1^{t-k}+A_2G_2^{t-k}+\cdots+A_{p'}G_{p'}^{t-k})=0 \tag{A4.1.8}$$

现在考虑

$$\begin{aligned}\varphi(B)&=(1-\varphi_1B-\varphi_2B^2-\cdots-\varphi_{p'}B^{p'})G_i^{t-k}\\&=G_i^{t-k-p'}(G_i^{p'}-\varphi_1G_i^{p'-1}-\cdots-\varphi_{p'})\end{aligned}$$

我们要证明对每个 i 有 $\varphi(B)G_i^{t-k}$ 为零。只须证明

$$G_i^{p'}-\varphi_1G_i^{p'-1}-\cdots-\varphi_{p'}=0$$

即 $B=1/G_i$ 是 $\varphi(B)=0$ 的根。现在我们来证明。(A4.16) 蕴含 $\varphi(B)=0$ 的根是 $B=1/G_i$，由此得到对每个 $i,\varphi(B)G_i^{t-k}$ 为零，因此 (A4.1.8) 成立，从而证明了 (A4.1.7) 是 (A4.1.5) 的通解。

为了直接推出 (A4.1.7)，我们考虑二阶方程的特殊情形

$$(1-G_1B)(1-G_2B)z_t=0$$

我们可将该式写为

$$(1-G_1B)y_t=0 \tag{A4.1.9}$$

其中

$$y_t=(1-G_2B)z_t \tag{A4.1.10}$$

那么，(A4.1.9) 蕴含

$$y_t=G_1y_{t-1}=G_1^2y_{t-2}=\cdots=G_1^{t-k}y_k$$

于是有

$$y_t=D_1G_1^{t-k}$$

这里，$D_1 = y_k$ 是由初始值 y_k 确定的常数。因此 (A4.1.10) 可以写成

$$
\begin{aligned}
z_t &= G_2 z_{t-1} + D_1 G_1^{t-k} \\
&= G_2(G_2 z_{t-2} + D_1 G_1^{t-k-1}) + D_1 G_1^{t-k} \\
&\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
&= G_2^{t-k} z_k + D_1(G_1^{t-k} + G_2 G_1^{t-k-1} + \cdots + G_2^{t-k-1} G_1) \\
&= G_2^{t-k} z_k + \frac{D_1}{1 - G_2/G_1}(G_1^{t-k} - G_2^{t-k}) \\
&= A_1 G_1^{t-k} + A_2 G_2^{t-k} \qquad \text{(A4.1.11)}
\end{aligned}
$$

这里 A_1, A_2 是由序列初始值确定的常数。将前述论证加以推广，可以证明当 $\varphi(B) = 0$ 具有非重根时，(A4.1.5) 的通解由 (A4.1.7) 给出。

重根　设 $\varphi(B) = 0$ 有 d 个相等的根 G_0^{-1}，因此 $\varphi(B)$ 包含因子 $(1 - G_0 B)^d$ 。特别，考虑二阶差分方程的解 (A4.1.11)，这时 G_1 和 G_2 都等于 G_0 。那么，(A4.1.11) 退化为

$$
z_t = G_0^{t-k} z_k + D_1 G_0^{t-k}(t-k)
$$

或

$$
z_t = [A_0 + A_1(t-k)] G_0^{t-k}
$$

一般，如果存在 d 重根 G_0，将其直接代入 (A4.1.5) 即可验证通解为

$$
z_t = [A_0 + A_1(t-k) + A_2(t-k)^2 + \cdots + A_{d-1}(t-k)^{d-1}] G_0^{t-k} \qquad \text{(A4.1.12)}
$$

特别，重根 G_0 都等于 1(例如 IMA$(0, d, q)$ 过程) 的情形，通解为

$$
z_t = A_0 + A_1(t-k) + \cdots + A_{d-1}(t-k)^{d-1} \qquad \text{(A4.1.13)}
$$

这是 $t-k$ 的 $d-1$ 阶多项式。

对一般情形，当 $\varphi(B)$ 的因式分解为

$$(1-G_1B)(1-G_2B)\cdots(1-G_pB)(1-G_0B)^d$$

余函数便是

$$C_k(t-k)=G_0^{t-k}\sum_{j=0}^{d-1}A_j(t-k)^j+\sum_{i=1}^{p}D_iG_i^{t-k} \qquad \text{(A4.1.14)}$$

因此，一般情况，余函数是由衰减指数项 G^{t-k}、多项式项 $(t-k)^j$、衰减正弦项 $D^{t-k}\sin(2\pi f_0t+F)$，以及这些函数的组合混合生成的。

“特解”的计算　现在我们来证明，满足方程

$$\varphi(B)I_k(t-k)=\theta(B)a_t \qquad (t-k>q) \qquad \text{(A4.1.15)}$$

的特解 $I_k(s-k)$，其定义如下：

$$\begin{aligned} I_k(s-k)&=0 \qquad (s<k)\\ I_k(1)&=a_{k+1}\\ I_k(2)&=a_{k+2}+\psi_1a_{k+1}\\ \vdots\quad &\quad \vdots\quad\vdots\quad\vdots\\ I_k(t-k)&=a_t+\psi_1a_{t-1}+\psi_2a_{t-2}+\cdots+\psi_{t-k-1}a_{k+1} \qquad (t>k) \end{aligned} \qquad \text{(A4.1.16)}$$

这里，ψ 权就是在模型的 (4.2.3) 表示式中出现的。因此，ψ 权满足

$$\varphi(B)\psi(B)a_t=\theta(B)a_t \qquad \text{(A4.1.17)}$$

现在可将 (A4.1.17) 的左边写成如下形式：

$$\begin{array}{ll|l} & a_t+\psi_1a_{t-1}+\psi_2a_{t-2}+\cdots+\psi_{t-k-1}a_{k+1} & +\psi_{t-k}a_k+\cdots\\ -\varphi_1(& \quad a_{t-1}+\psi_1a_{t-2}+\cdots+\psi_{t-k-2}a_{k+1} & +\psi_{t-k-1}a_k+\cdots\\ -\varphi_2(& \quad\cdots\quad\cdots\quad\cdots\quad\cdots & \cdots\quad\cdots\quad\cdots\\ \vdots & \quad\vdots\quad\vdots\quad\vdots\quad\vdots & \vdots\quad\vdots\quad\vdots\\ -\varphi_{p'}(& \qquad a_{t-p'}+\cdots+\psi_{t-k-p'}a_{k+1} & +\phi_{t-k-p'}a_k+\cdots \end{array} \qquad \text{(A4.1.18)}$$

由于 (A4.1.17) 的右边是

$$a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

于是得到前面矩阵中的前 $q+1$ 列之和分别等于 $a_t, -\theta_1 a_{t-1}, \cdots,$ $-\theta_q a_{t-q}$. (A4.1.15) 式的左边等于矩阵中前 $t-k$ 列之和，即垂直线左边的那些项，这里， $I_k(s-k)$ 由 (A4.1.16) 给出。因此，如果 $t-k<q$, 那么，垂直线划在 $q+1$ 列之后，这时垂直线之前所有项之和等于 $\theta(B)a_t$. 这表明 (A4.1.16) 是差分方程的一个"特解".

例：考虑 IMA(0,1,1) 过程

$$z_t - z_{t-1} = a_t - \theta a_{t-1} \tag{A4.1.19}$$

其中 $\psi_j = 1-\theta$, 对 $j > 1$. 于是

$$\begin{aligned} I_k(0) &= 0 \\ I_k(1) &= a_{k+1} \\ \vdots \quad & \quad \vdots \\ I_k(t-k) &= a_t + (1-\theta)\sum_{j=1}^{t-k-1} a_{t-j} \qquad (t-k>1) \end{aligned} \tag{A4.1.20}$$

如果 $z_k = I_k(t-k)$ 是 (A4.1.19) 的解，那么

$$I_k(t-k) - I_k(t-k-1) = a_t - \theta a_{t-1}$$

容易验证, 当 $t-k=1$ 时 (A4.1.20) 不满足上式, 只有当 $t-k>1$(也就是 $t-k>q$ 时， (A4.1.20) 才满足上式。

A4.2 具有确定性偏差的 IMA(0,1,1) 过程

一般模型 $\phi(B)\nabla^d z_t = \theta_0 + \theta(B)\epsilon_t$ 可以写成

$$\phi(B)\nabla^d z_t = \theta(B)\epsilon_t$$

其中冲击具有非零均值 $\xi = \theta_0/(1-\theta_1-\cdots-\theta_q)$. 比如, IMA(0,1,1) 模型可表示成

$$\nabla z_t = (1-\theta B)\epsilon_t$$

这里 $E[\epsilon_t] = \xi = \theta_0/(1-\theta)$ 。例如，用加热器按照固定的速率供热，这时反应器出口的温度 z_t 就可以用这种形式表式. 现在，若

$$\epsilon_t = \xi + a_t \tag{A4.2.1}$$

其中 a_t 是零均值白噪声，于是相对于时间原点 k 模型的求和形式为

$$z_t = b_0^{(k)} + \lambda \sum_{j=1}^{t-k-1} \epsilon_{t-j} + \epsilon_t \tag{A4.2.2}$$

式中 $\lambda = 1-\theta$ 。把 (A4.2.1) 代入 (A4.2.2), 模型可用 a 表示为

$$z_t = b_0^{(k)} + \lambda\xi(t-k-1) + \xi + \lambda \sum_{j=1}^{t-k-1} a_{t-j} + a_t \tag{A4.2.3}$$

由此我们可以看出，z_t 含有由 $\lambda\xi(t-k-1)$ 给出的确定性的斜率或变化态势，确定的线性趋势的斜率等于 $\lambda\xi = \theta_0$ 。此外，如果我们把过程在时刻 $t-1$ 的“水平”记为 l_{t-1}, 这里

$$z_t = l_{t-1} + a_t$$

我们看到，从时刻 $t-1$ 到 t 水平的改变是依照如下方式:

$$l_t = l_{t-1} + \lambda\xi + \lambda a_t$$

因此，水平分量的变化包含一个确定性分量 $\lambda\xi = \theta_0$, 同时还有一个随机分量 λa_t 。

A4.3 带有附加噪声的ARIMA过程

在本附录中，我们考虑附加噪声 (例如，量测误差) 对一般 ARIMA 过程的影响。

A4.3.1 两个独立滑动平均过程之和

为了对下文讨论作必要的准备，我们来考虑随机过程 w_t, 它是阶数分别为 q_1 和 q_2 的两个独立滑动平均过程之和。即是

$$w_t = \theta_1(B)a_t + \theta_2(B)b_t \tag{A4.3.1}$$

其中 $\theta_1(B)$ 和 $\theta_2(B)$ 分别是 B 的 q_1 阶和 q_2 阶多项式，白噪声过程 a_t 和 b_t 都具有零均值且相互独立。设 $q = \max(q_1, q_2)$, 则显然 w_t 的自相关函数 γ_j 当 $j > q$ 时必为零。由此可知，可以把 w_t 表示成 q 阶的单一滑动平均过程:

$$w_t = \theta_3(B)u_t \tag{A4.3.2}$$

这里， u_t 是均值为零的白噪声过程。因此，两个独立滑动平均过程之和是另一个滑动平均过程，相应阶数与分量过程中阶数较高者一样。

A4.3.2 附加噪声对一般模型的影响

相关噪声　考虑 (p,d,q) 阶的一般非平稳模型

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{A4.3.3}$$

假设我们不能观测到 z_t 本身，只能观测到 $Z_t = z_t + b_t$, 这里， b_t 表示有关的某种附加噪声 (如测量误差) . 我们希望确定观测过程 Z_t 的性质. 一般，我们有

$$\phi(B)\nabla^d Z_t = \theta(B)a_t + \phi(B)\nabla^d b_t$$

若噪声遵从平稳的 $(p_1, 0, q_1)$ 阶 ARMA 过程，即

$$\phi_1(B)b_t = \theta_1(B)\alpha_t \tag{A4.3.4}$$

其中 α_t 是与 a_t 独立的白噪声过程，那么

$$\underbrace{\phi_1(B)\phi(B)\nabla^d Z_t}_{p_1+p+d} = \underbrace{\varphi_1(B)\theta(B)a_t}_{p_1+q} + \underbrace{\phi(B)\theta_1(B)\nabla^d \alpha_t}_{p+q_1+d} \tag{A4.3.5}$$

花括号下方的值标明了 B 的各个多项式的阶数. 现在 (A4.3.5) 的右边和 (A4.3.1) 的形式是一样的. 令 $Q = p_1 + p$, 而 Q 是 $(p_1 + q)$ 和 $(p + q_1 + d)$ 中较大者. 于是我们可以写成

$$\phi_2(B)\nabla^d Z_t = \theta_2(B)u_t$$

其中 u_t 是白噪声过程，而 Z_t 可看作是 (P, d, Q) 阶 ARIMA 过程.

附加白噪声 在一些应用中确实可能出现附加噪声为白噪声的情形，即在 (A4.3.4) 中 $\phi_1(B) = \theta_1(B) = 1$, 我们得到

$$\phi(B)\nabla^d Z_t = \theta_2(B)u_t \tag{A4.3.6}$$

其中

$$\theta_2(B)w_t = \theta(B)a_t + \phi(B)\nabla^d b_t$$

这里，模型为 (p, d, Q) 阶，而 Q 是 q 和 $(p + d)$ 中的较大者. 如果 $p + d < q$, 则带有误差的过程与原过程的阶数一样. 附加噪声的影响只是改变 θ 的值 (而不改变 ϕ) .

附加白噪声对求和滑动平均过程的影响 特别，一个 $(0, d, q)$ 阶附加白噪声的 IMA 过程，当 $d < q$ 时，仍然是一个 $(0, d, q)$ 阶 IMA 过程，否则将会成为 $(0, d, d)$ 阶的 IMA 过程. 在以上两种情形，由于噪声的添加，过程的参数都将改变. 通过对附加噪声过程和单纯 IMA 过程二者的 d 阶差分求的自协方差，并令其相

等，我们能够确定参数改变的特性。以上处理过程将用一个例子来说明。

A4.3.3 附加白噪声 IMA(0,1,1) 过程的例子

若 $Z_t = z_t + b_t$，考虑该过程的性质，这里

$$z_t = z_{t-1} - (1-\lambda)a_{t-1} + a_t \tag{A4.3.7}$$

其中，b_t 和 a_t 是互相独立的白噪声过程。过程 Z_t 的一阶差分为 $W_t = Z_t - Z_{t-1}$ 并有

$$W_t = [1-(1-\lambda)B]a_t + (1-B)b_t \tag{A4.3.8}$$

一阶差分 W_t 的自协方差为

$$\begin{aligned} \gamma_0 &= \sigma_a^2[1+(1-\lambda)^2] + 2\sigma_b^2 \\ \gamma_1 &= -\sigma_a^2(1-\lambda) - \sigma_b^2 \\ \gamma_j &= 0 \qquad (j > 2) \end{aligned} \tag{A4.3.9}$$

滞后超过 1 的 γ_j 为零，这证实了我们先前的预料：附加白噪声过程就是一个 (0,1,1) 阶的 IMA 过程。为了表示出附加噪声 IMA 过程的参数，我们假定该过程可写为：

$$Z_t = Z_{t-1} - (1-\Lambda)u_{t-1} + u_t \tag{A4.3.10}$$

其中 u_t 是噪声过程。过程 (A4.3.10) 的一阶差分为 $W_t = Z_t - Z_{t-1}$，其协方差是

$$\begin{aligned} \gamma_0 &= \sigma_u^2[1+(1-\Lambda)^2] \\ \gamma_1 &= -\sigma_u^2(1-\Lambda) \\ \gamma_j &= 0 \qquad (j > 2) \end{aligned} \tag{A4.3.11}$$

(A4.3.9) 和 (A4.3.11) 是相等的，于是我们可将 Λ 和 σ_u^2 解出，即是

$$\begin{aligned} \frac{\Lambda^2}{1-\Lambda} &= \frac{\lambda^2}{1-\lambda+\sigma_b^2/\sigma_a^2} \\ \sigma_u^2 &= \sigma_a^2\frac{\lambda^2}{\Lambda^2} \end{aligned} \tag{A4.3.12}$$

譬如，原序列 $\lambda = 0.5$ 及 $\sigma_b^2 = \sigma_a^2$, 那么 $\Lambda = 0.33$, $\sigma_u^2 = 2.25\sigma_a^2$ 。

A4.3.4 IMA(0,1,1) 过程和随机游动之间的关系

过程

$$z_t = z_{t-1} + a_t \tag{A4.3.13}$$

是一个 $\lambda = 1$ 的 IMA(0,1,1) 过程，有时称之为随机游动。如果 a_t 取为在 t 时刻向前或向后的步数，那么 z_t 就表示在 t 时刻游动的位置。任何一个 IMA(0,1,1) 过程都可以认为是隐含于白噪声 b_t 中的随机游动，这里，它和与随机游动联系的冲击 a_t 不相关。若带噪声的过程是 $Z_t = z_t + b_t$, 其中 z_t 由 (A4.3.13) 定义，于是利用 (A4.3.12) 有

$$Z_t = Z_{t-1} - (1-\Lambda)u_{t-1} + u_t$$

这里

$$\frac{\Lambda^2}{1-\Lambda} = \frac{\sigma_a^2}{\sigma_b^2}, \qquad \sigma_u^2 = \frac{\sigma_a^2}{\Lambda^2} \tag{A4.3.14}$$

A4.3.5 附加相关噪声一般过程的协方差函数

假定基本过程为 (p,d,q) 阶 AIRMA 过程

$$\phi(B)\nabla^d z_t = \theta(B)a_t$$

而 $Z_t = z_t + b_t$ 是观测到的过程， b_t 是协方差函数为 $\gamma_j(b)$ 的平稳过程，b_t 与过程 a_t 独立，从而也与 z_t 独立。假设 $w_t = \nabla^d z_t = \phi^{-1}(B)\theta(B)a_t$ 及 $W_t = \nabla^d Z_t$ 。我们需要求出 W_t 的自协方差函数。现在有

$$\nabla^d(Z_t - b_t) = \phi^{-1}(B)\theta(B)a_t$$
$$W_t = w_t + v_t$$

其中

$$v_t = \nabla^d b_t = (1-B)^d b_t$$

因此

$$\begin{aligned}
\gamma_j(W) &= \gamma_j(w) + \gamma_j(v) \\
\gamma_j(v) &= (1-B)^d(1-F)^d\gamma_j(b) \\
&= (-1)^d(1-B)^{2d}\gamma_{j+d}(b)
\end{aligned}$$

及

$$\gamma_j(W) = \gamma_j(w) + (-1)^d(1-B)^{2d}\gamma_{j+d}(b) \qquad \text{(A4.3.15)}$$

例如，假设 IMA(0,1,1) 过程定义为 $w_t = \nabla z_t = (1-\theta B)a_t$，相关噪声 b_t 被加到过程 w_t 上，那么，带噪声过程一阶差分的协方差是

$$\begin{aligned}
\gamma_0(W) &= \sigma_a^2(1+\theta^2) + 2[\gamma_0(b) - \gamma_1(b)] \\
\gamma_1(W) &= -\sigma_a^2\theta + [2\gamma_1(b) - \gamma_0(b) - \gamma_2(b)] \\
\gamma_j(W) &= [2\gamma_j(b) - \gamma_{j-1}(b) - \gamma_{J+1}(b)] \qquad (j > 2)
\end{aligned}$$

特别，若 b_t 是一阶自回归过程，则

$$\begin{aligned}
b_t &= \phi b_{t-1} + \alpha_t \\
\gamma_0(W) &= \sigma_a^2(1+\theta^2) + 2\sigma_b^2(1-\phi) \\
\gamma_1(W) &= -\sigma_a^2\theta - \sigma_b^2(1-\phi)^2 \\
\gamma_j(W) &= -\sigma_b^2\phi^{j-1}(1-\phi)^2 \qquad (j > 2)
\end{aligned}$$

事实上，在这种情形下，(A4.3.5) 产生的带噪声过程 $Z_t = z_t + b_t$ 可定义如下

$$(1-\phi B)\nabla Z_t = (1-\phi B)(1-\theta B)a_t + (1-B)\alpha_t$$

模型的阶为 (1,1,2)．

5

预　报

第 4 章中我们考虑了 ARIMA 模型的一些性质，在本章中将要说明如何利用有关性质去对观测值的未来进行预报. 在本书的第二部分我们将考虑对实际数据的模型拟合问题，因此，目前我们是把模型当作完全已知而着手研究. 须知，如果用以拟合模型的数据点不是很少，那么参数估计误差对预报的影响不会很大.

在本章中我们考虑非季节时间序列，季节时间序列的预报在第 9 章中讨论. 我们将要说明，如何从模型的差分形式直接得出最小均方误差预报，而进一步的递推计算可得到预报概率限. 应该指出，对于具体计算预报而言，通过差分方程的处理办法是最简便最好的，但是为了深入理解预报的性质，我们还要从其它角度加以考虑.

5.1　最小均方误差预报及其性质

在 4.2 节中我们所讨论的一般 ARIMA 模型的显式为

$$\varphi(B)z_t = \theta(B)a_t \tag{5.1.1}$$

其中 $\varphi(B) = \phi(B)\nabla^d$. 在节 4.2 中我们考虑了 ARIMA 模型的三种显式表示，我们仍从这三种表示入手，因为每一种表示都可从不同的侧面对预报问题加以说明.

若当前时刻为 t, 我们将要涉及的是对值 $z_{t+l}, l > 0$ 的预报.

这种预报称为以 t 为原点，提前期 (或步长) 为 l 的预报。我们来归纳一下节 4.2 的结果，只是这里用 $t+l$ 来代替 t, 用 t 代替 k 。

模型的三种显式表示　由过程 (5.11) 产生的观测 z_{t+l} 可以表示如下

1. 直接用差分方程表示为

$$z_{t+l}=\varphi_1 z_{t+l-1}+\cdots+\varphi_{p+d}z_{t+l-p-d}-\theta_1 a_{t+l-1}-\cdots-\theta_{t+l-q}+a_{t+l} \tag{5.1.2}$$

2. 表示为当前和过去冲击 a_j 的无穷加权和：

$$z_{t+l}=\sum_{j=0}^{\infty}\psi_j a_{t+l-j} \tag{5.1.3}$$

这里 $\psi_0=1$. 在下式中：

$$\varphi(B)(1+\psi_1 B+\psi_2 B^2+\cdots)=\theta(B) \tag{5.1.4}$$

令系数相等，与 (4.2.5) 式一样，可以得到 ψ 权。l 为正时等价于参考原点 $K<t$ 的情形，模型可以写成余函数形式

$$\begin{aligned} z_{t+l} &= a_{t+l}+\psi_1 a_{t+l-1}+\cdots+\psi_{l-1}a_{t+1} \\ &\qquad +\psi_l a_t+\cdots+\psi_{t+l-k-1}a_{k+1}+C_k(t+l-k) \\ &= a_{t+l}+\psi_1 a_{t+l-1}+\cdots+\psi_{l-1}a_{t+1}+C_t(l) \end{aligned} \tag{5.1.5}$$

其中 $C_k(t+l-k)$ 是相对于过程有穷原点 k 的余函数。与 (4.2.19) 对照，我们可以把相对于预报原点 t 的余函数表示成 $C_t(l)=C_k(t+l-k)+\psi_l a_t+\psi_{l+1}a_{t-1}+\cdots+\psi_{t+l-k-1}a_{k+1}$ 。 $C_t(l)$ 可形式上表示为截断的无穷和：

$$C_t(l)=\sum_{j=l}^{\infty}\psi_j a_{t+l-j} \tag{5.1.6}$$

3. 表示为过去观测的无穷加权和再加上一个随机冲击

$$z_{t+l}=\sum_{j=1}^{\infty}\pi_j z_{t+l-j}+a_{t+l} \tag{5.1.7}$$

即若 $d > 1$, 有

$$\bar{z}_{t+l-1}(\pi) = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} \tag{5.1.8}$$

由于 $\sum\limits_{j=1}^{\infty}\pi_j = 1$, 上式是加权平均。

与 (4.2.22) 一样, π 权可由下式得到

$$\varphi(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots)\theta(B) \tag{5.1.9}$$

5.1.1 最小均方误差预报的推导

现假定当前时刻为 t, 我们要做出 z_{t+l} 的预报值 $z_t(l)$, 它是当前和历史观测值 $z_t, z_{t-1}, z_{t-2}, \cdots$ 的线性函数, 因而也是现在和过去冲击 $a_t, a_{t-1}, a_{t-2}, \cdots$ 的线性函数。

假设最佳预报是

$$\hat{z}_t(l) = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \cdots$$

权系数 $\psi_l^*, \psi_{l+1}^*, \cdots$ 是待定的. 那么利用 (5.1.3) 预报的均方误差是

$$E[z_{t+l} - \hat{z}_t(l)]^2 = (1 + \psi_1^2 + \cdots + \psi_{l_1}^2)\sigma_a^2 + \sum_{j=0}^{\infty}(\psi_{l+j} - \psi_{l+j}^*)^2\sigma_a^2 \tag{5.1.10}$$

令 $\psi_{l+j}^* = \psi_{l+j}$ 可使上式极小化, 该结论是预报理论中更一般结果的特殊情形, 可参见 Wold[207],Kolmogoroff[130],[131],[132],Wiener [200] 和 Whittle[198] . 于是我们有

$$\begin{aligned} z_{t+l} &= (a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1}) \\ &\quad + (\psi_l a_t + \psi_{l+1} a_{t-1} + \cdots) \qquad (5.1.11) \\ &= e_t(l) + \hat{z}_t(l) \qquad (5.1.12) \end{aligned}$$

其中 $e_t(l)$ 是提前 l 期预报值 $\hat{z}_t(l)$ 的误差。

我们应注意一些重要事实。如前所述，用 $E[z_{t+l} \mid z_t, z_{t-1}, \cdots]$ 表示直到 t 时刻之前全部 z 作为已知时 z_{t+l} 的条件期望，简记为 $E_t[z_{t+l}]$，我们将假定 a_t 是独立随机变量序列。以下结论成立：

1. $E[a_{t+j} \mid z_t, z_{t-1}, \cdots] = 0$，$j > 0$，且由 (5.1.3) 得

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots = E_t[z_{t+l}] \tag{5.1.13}$$

因此，在原点 t 的提前 l 期的最小均方误差预报就是 z_{t+l} 在时刻 t 的条件期望。如果把 $\hat{z}_t(l)$ 看作是当 t 固定时 l 的函数，我们称之为关于原点 t 的预报函数。我们注意到，为了使条件期望 $E_t[z_{t+l}]$(它恒等于最小均方误差预报) 与最小均方线性预报一致，在模型 (5.1.1) 中对于随机冲击 a_t 的最低要求就是 $E_t[(a_{t+j})] = 0, j > 0$ 。这种情形对某些本质为非线形的过程可能会不成立，有关问题的研究近来受到极大关注 (例如，参见 [160],[188]) 。事实上，这类过程可能具有形如 (5.1.1) 的线性表示，冲击 a_t 并不是独立而只是不相关的，那么最佳预报 $E_t[z_{t+l}]$ 可以与由 (5.1.11) 得到的最佳线性预报 $\hat{z}_t(l)$ 不相符。

对于 $d > 1$ 的非平稳情形，前文中的 (5.1.13) 式只是形式上基于无穷冲击表示式 (5.1.3)，我们也可以形式上由模型 (5.1.5) 的截断形式推出 (5.1.13) 。为此，我们假设相对于初始固定原点 $k < t$ 的余函数 $C_k(t-k)$(在时刻 k 已确定) 与整个冲击序列 $a_{k+i}, i > 0$ 是独立的。因此，对 (5.1.5) 中第一方程的右边取条件期望，我们得到

$$\begin{aligned} E_t[z_{t+l}] = \hat{z}_t(l) = \psi_l z_t + \psi_{l+1} a_{t-1} + \cdots \\ + \psi_{t+l-k-1} a_{k+1} + C_k(t+l-k) \end{aligned}$$

这就是在原点 t 的最小均方误差预报。该式也可用 (5.1.11) 右边第二项作形式上的表示。

2. 提前 l 期的预报误差是

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} \tag{5.1.14}$$

由于

$$E_t[e_t(l)] = 0 \tag{5.1.15}$$

故预报是无偏的。且预报误差的方差为

$$V(l) = \text{var}[e_t(l)] = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 \tag{5.1.16}$$

3. 容易证明，不仅 $\hat{z}_t(l)$ 是 z_{t+l} 的最小均方误差预报，且预报值的任意线性函数 $\sum_{l=1}^{L} w_l \hat{z}_t(l)$ 也是未来观测值相应函数 $\sum_{l=1}^{L} w_l z_{t+l}$ 的最小均方误差预报。例如，假设由月度数据我们利用 (5.1.13) 已得出了未来第 1,2,3 个月产品销售量的预报值 $\hat{z}_t(1), \hat{z}_t(2), \hat{z}_t(3)$，那么，下一季度销售量 $z_{t+1} + z_{t+2} + z_{t+3}$ 的最小均方误差预报就是 $\hat{z}_t(1) + \hat{z}_t(2) + \hat{z}_t(3)$。

4. 提前 1 期的预报误差就是冲击。利用 (5.1.14)，提前 1 期的预报误差是：

$$e_t(1) = z_{t+1} - \hat{z}(1) = a_{t+1} \tag{5.1.17}$$

因此，在此之前，对于生成过程的冲量 a_t 我们还只是作为一组独立的随机变量或冲击而引入，现在证明了它就是提前 1 期预报的误差。由此得知，对于最小均方误差预报而言，提前 1 期的预报误差必然是不相关的。这一点是很明显的，因为如果提前 1 期预报误差是相关的，那么在一定程度上预报误差 a_{t+1} 就可以用已有的预报误差 $a_t, a_{t-1}, a_{t-2}, \ldots$ 去预报，若由此得到的预报为 $\hat{a}_{t+1}$，那么 $\hat{z}_t(1) + \hat{a}_{t+1}$ 就是比 $\hat{z}_t(1)$ 更好的 z_{t+1} 的最佳预报。

5. 预报误差之间的相关。尽管提前一期最优预报误差是不相关的，但是对于较长的预报期，预报误差一般就是相关的了。在节 A5.1.1 中我们推导了从不同原点 t 和 $t-j$ 出发，都是 l 期预报的误差 $e_t(l)$ 和 $e_{t-j}(l)$ 之间相关的一般表达式。同样，从同一原点出发，做不同期数预报的预报误差 $e_t(l)$ 和 $e_t(l+j)$ 之间也是相关的。其后果是预报函数常常会存在趋势性，即相对于最终得到的观测序列来说，预报值全都位于其上方或下方。在节 A5.1.2 中我们给出从相同原点出发所作预报误差 $e_t(l)$ 和 $e_t(l+j)$ 之间相关的一般表达式。

5.1.2 预报的三种基本形式

我们已经看到，提前 l 期最小均方误差预报 $\hat{z}_t(l)$ 是 z_{t+l} 在原点 t 的条件期望 $E_t[z_{t+l}]$，利用这一事实，我们可以写出三种不同形式的预报表达式，这对应于在本节开始所归纳的表示模型的三种方式. 为了记号方便，我们暂时约定用方括号表示在 t 的条件期望，即

$$[a_{t+l}] = E_t[a_{t+l}] \qquad [z_{t+l}] = E_t[z_{t+l}]$$

对于 $k > 0$，表示预报的三种不同形式为:

由差分方程得到的预报　对 (5.1.2) 在时刻 t 取条件期望，我们得到

$$\begin{aligned}[z_{t+l}] = \hat{z}_t(l) = \varphi_1[z_{t+l-1}] + \cdots + \varphi_{p+d}[z_{t+l-p-d}] \\ - \theta_1[a_{t+l-1}] \cdots - \theta_q[a_{t+l-q}] + [a_{t+l}]\end{aligned} \tag{5.1.18}$$

求和形式表示的预报　利用 (5.1.3) 我们有

$$\begin{aligned}[z_{t+l}] = \hat{z}_t(l) = [a_{t+l}] + \psi_1[a_{t+l-1}] + \cdots + \psi_{l-1}[a_{t+l}] \\ + \psi_l[a_t] + \psi_{l+1}[a_{t-1}] + \cdots\end{aligned} \tag{5.1.19}$$

这是前文中已见到的 (5.1.13) 形式. 另外，利用模型 (5.1.5) 的截断形式，我们有

$$\begin{aligned}[z_{t+l}] = \hat{z}_t(l) &= [a_{t+l}] + \psi_1[a_{t+l-1}] + \cdots \\ &\quad + \psi_{t+l-k-1}[a_{k+1}] + C_k(t+l-k) \\ &= [a_{t+l}] + \psi_1[a_{t+l-1}] + \cdots + \psi_{l-1}[a_{t+1}] + C_t(l)\end{aligned} \tag{5.1.20}$$

这里 $C_t(l)$ 是在原点 t 的余函数.

预报作为过去观测值的加权平均,以及利用从同一原点的过去几期预报值所做的预报　最后，在 (5.1.7) 中取条件期望可得到

$$[z_{t+l}] = \hat{z}_t(l) = \sum_{j=1}^{\infty} \pi[z_{t+l-j}] + [a_{t+l}] \qquad (5.1.21)$$

应该注意到，最小均方误差预报是用如下的条件方差所定义的：

$$[z_{t+l}] = E_t[z_{t+l}] = E[z_{t+l} \mid z_t, z_{t-1}, \cdots]$$

在理论上，这需要追溯到无穷过去的所有 z 的信息，但是，我们赋予一般 ARIMA 模型的可逆性条件保证了 (5.1.21) 中的 π 权构成收敛级数。因此，为了在给定精度下计算预报值，可取某个 k，当 $j > k$ 时上述和式对 z_{t-j} 的依赖性可以忽略。在实际中，π 权通常收敛得较快，因此在计算时无论采用哪一种形式的模型，只需要适当长的序列 $z_t, z_{t-1}, \cdots z_{t-k}$ 就能使预报达到足够的精度。我们所讨论的方法稍加改进就可以计算精确的有穷样本预报 $E[z_{t+l} \mid z_t, z_{t-l}, \cdots, z_1]$，这是基于有限长度数据 $z_t, z_{t-1}, \cdots, z_1$ 的预报。

为了计算在表达式 (5.1.18) 到 (5.1.21) 中出现的条件期望。我们注意到，若 j 是非负整数则有

$$\begin{aligned}
&[z_{t-j}] = E_t[z_{t-j}] = z_{t-j} && (j = 0, 1, 2, \cdots) \\
&[z_{t+j}] = E_t[z_{t+j}] = \hat{z}_t(j) && (j = 1, 2, \cdots) \\
&[a_{t-j}] = E_t[a_{t-j}] = a_{t-j} && \\
&\qquad\quad = z_{t-j} - \hat{z}_{t-j-1}(1) && (j = 0, 1, 2, \cdots) \\
&[a_{t+j}] = E_t[a_{t+j}] = 0 && (j = 1, 2, \cdots)
\end{aligned} \qquad (5.1.22)$$

因此，为了得出预报值 $\hat{z}_t(l)$，我们可用前文所述三种显式表示的任何一种写出关于 z_{t+l} 的模型，并按照下面规则处理右边各项：

1. $z_{t-j} (j = 0, 1, 2, \cdots)$ 是在原点时刻 t 已经发生的，故保持不变。

2. $z_{t+j} (j = 1, 2, \cdots)$ 是尚未发生的，用在原点 t 的预报值 $\hat{z}_t(j)$ 来代替。

3. $a_{t-j} (j = 0, 1, 2, \cdots)$ 是已发生的，用 $z_{t-j} - \hat{z}_{t-j-1}(1)$ 得出。

4. $a_{t+j}(j=1,2,\cdots)$ 是尚未发生的，用 0 代替.

对常规计算来说，直接使用差分方程形式 (5.1.18) 最简单，因此预报值可按下面的显式递推算出：

$$\hat{z}_t(l)=\sum_{j=1}^{p+d}\varphi_j\hat{z}_t(l-j)-\sum_{j=1}^{q}\theta_j a_{t+l-j}$$

这里 $\hat{z}_t(-j)=[z_{t-j}]$, 当 $j>0$ 时表示观测值，当预报提前期 $l>q$ 时，上式中的滑动平均项不存在.

例：用差分方程形式进行预报　在第七章中将说明，序列 C 可用下面模型近似表示：

$$(1-0.8B)(1-B)z_{t+1}=a_{t+1}$$

即

$$(1-1.8B+0.8B^2)z_{t+1}=a_{t+1}$$

或

$$z_{t+l}=1.8z_{t+l-1}-0.8z_{t+l-2}+a_{t+l}$$

原点 t 的预报是

$$\begin{aligned}\hat{z}_t(1)&=1.8z_t-0.8z_{t-1}\\ \hat{z}_t(2)&=1.8\hat{z}_t(1)-0.8z_t\\ \hat{z}_t(l)&=1.8\hat{z}_t(l-1)-0.8\hat{z}_t(l-2)\qquad(l=3,4,\cdots)\end{aligned}\tag{5.1.23}$$

可以看出，预报值可依次用 $\hat{z}_t(1),\hat{z}_t(2),\cdots$ 递推产生. 在前例中，恰好没有滑动平均项，如果存在滑动平均项也不会造成额外的困难. 在本章稍后，我们要考虑在控制问题中出现的一个序列，在 $t+l$ 时刻相应的模型是

$$\nabla^2z_{t+l}=(1-0.9B+0.5B^2)a_{t+l}$$

于是

$$
\begin{aligned}
z_{t+l} &= 2z_{t+l-1} - z_{t+l-2} + a_{t+l} - 0.9a_{t+l-1} + 0.5a_{t+l-2} \\
\hat{z}_t(1) &= 2z_t - z_{t-1} - 0.9a_t + 0.5a_{t-1} \\
\hat{z}_t(2) &= 2\hat{z}_t(1) - z_t + 0.5a_t \\
\hat{z}_t(l) &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \qquad (l = 3, 4, \cdots)
\end{aligned}
$$

在以上表达式中我们应当记住: $a_t = z_t - \hat{z}_{t-1}(1), a_{t-1} = z_{t-1} - \hat{z}_{t-2}(1)$, 且在预报过程刚开始时, 可令未知的 a 等于它们的无条件期望值 (即为 0) 。

按照约定, 通常假设数据从时刻 $s = 1$ 开始可以使用, 利用模型的差分方程形式 (5.1.2) 进行递推计算所需要的 a_s 是

$$
\begin{aligned}
a_s &= z_s - \hat{z}_{s-1}(1) \\
&= z_s - \left(\sum_{j=1}^{p+d}\varphi_j z_{s-j} - \sum_{j=1}^{q}\theta_j a_{s-j}\right) \qquad (s = p + d + 1, \cdots, t)
\end{aligned}
$$

对于 $s < p + d + 1$, 取初值 a_s 为零. 利用另一种方法我可以得到更精确的预报形式, 即对刚开始预报时必须用到的 a_s 和 z_s 初值, 采用反向预报的技术加以确定. 这一技术的实质就是确定初值 a_s 和 z_s 关于有效数据段的条件期望, 以此作为从时刻 $s = 1$ 起的有限数据段. 这种技巧在第 7 章有关 ARIMA 模型参数估计内容中还要讨论. 运用反向预报技术获得 a 和 z 必需的初值, 在利用样本数据 $z_t, z_{t-1}, \cdots, z_1$ 的有限过去信息基础上, 我们最终得出了的最小均方误差预报 $E[z_{t+l} \mid z_t, z_{t-1}, \cdots, z_1]$ 。如果能够提供充分长的数据序列 $z_t, z_{t-1}, \cdots, z_1$, 那么, 两种不同的初值处理方法对预报值 $\hat{z}_t(l)$ 的影响是可以忽略的.

对于 ARI(p, d) 的特殊情形 (即 $q = 0$ 的 ARIMA 过程), 预报值不依赖于 a 或 z 的初值, 事实上, 从 (5.1.18) 看出, 预报值为

$$
\hat{z}_t(l) = \varphi_1\hat{z}_t(l-1) + \varphi_2\hat{z}_t(l-2) + \cdots + \varphi_{p+d}\hat{z}_t(l-p-d) \quad (l = 1, 2, \cdots)
$$

起始值是 $\hat{z}_t(1) = \varphi_1 z_t + \varphi_2 z_{t-1} + \cdots + \varphi_{p+d} z_{t+1-p-d}$ 。由此我们

看到，这一模型的任意 l 期预报 $\hat{z}_t(l)$ 只依赖于该过程最邻近的 $p+d$ 个观测值 $z_t, z_{t-1}, \cdots, z_{t+1-p-d}$，但当预报期较长就并非如此了．当然，并不能认为这时 a 的影响就不包含在预报值中．例如，在前文的例子中，$\hat{z}_t(3)$ 依赖于 $\hat{z}_t(2)$ 和 $\hat{z}_t(1)$，它同样也依赖于 a_t 和 a_{t-1}。

最后，我们注意到对于 $d>1$ 的 ARIMA 过程，另一种考察预报的方法可能是有用的．例如，对于 $d=1$ 我们有 $w_t=(1-B)z_t$ 或 $z_t=w_t+z_{t-1}$，这里，w_t 是平稳 $ARMA(p,q)$ 过程。因此过程 z 的预报可表示为

$$\hat{z}_t(l)=\hat{w}_t(l)+\hat{z}_t(l-1)=\hat{w}_t(l)+\hat{w}_t(l-1)+\cdots+\hat{w}_t(1)+z_t$$

其中，预报值 $\hat{w}_t(l)$ 是由 w 的 ARMA(p,q) 过程基于数据 $w_t, w_{t-1}, \cdots$ 得到。类似地，对于 $d=2$，由 $w_t=(1-B)^2z_t=z_t-2z_{t-1}+z_{t-2}$ 我们得到 $\hat{z}_t(l)=\hat{w}_t(l)+2\hat{z}_t(l-1)-\hat{z}_t(l-2)=\sum_{j=0}^{l-1}(j+1)\hat{w}_t(l-j)+z_t+l(z_t-z_{t-1})$，等等。

5.2 预报的计算和修正

5.2.1 进行预报的方便格式

经常会遇到这样的情形：我们需要作提前多期的向前预报，如提前 1 期，2 期，$\cdots, L$ 期。我们按照表 5.1 所示计算格式很容易由其它一些值把每期预报值算出，在表中给出的预报是以 $t=20$ 为原点对序列 C 做出的，预报期为 $l=1,2,3,\cdots,14$。对角线的排列形式使得每个预报值在对着被预报值的位置出现．如 $\hat{z}_{20}(6)=22.51$ 是在原点 20 对 z_{26} 做的预报，因此它的值对着 z_{26}。在 $t=21,22$ 等等的 z_t 的实际值用较大字体表示，为的是提醒大家注意：在进行预报时，这些值实际上是不能使用的。

利用 (5.1.23) 很容易进行计算。例如

$$\hat{z}_{20}(1)=(1.8\times 23.4)-(0.8\times 23.7)=23.16$$

$$\hat{z}_{20}(2)=(1.8\times 23.16)-(0.8\times 23.4)=22.97$$

表 5.1 预报的方便格式

				1	2	3	4	5	6	7	8	9	10	11	12	13	14
提前期 l→				1	2	3	4	5	6	7	8	9	10	11	12	13	14
系数 ψ_l→				1.80	2.44	2.95	3.36	3.69	3.95	4.16	4.33	4.46	4.57	4.65	4.72	4.78	4.82
预报特性 →				$\hat{z}_{t-1}(1)$	$\hat{z}_{t-2}(2)$	$\hat{z}_{t-3}(3)$	$\hat{z}_{t-4}(4)$	$\hat{z}_{t-5}(5)$	$\hat{z}_{t-6}(6)$	$\hat{z}_{t-7}(7)$	$\hat{z}_{t-8}(8)$	$\hat{z}_{t-9}(9)$	$\hat{z}_{t-10}(10)$	$\hat{z}_{t-11}(11)$	$\hat{z}_{t-12}(12)$	$\hat{z}_{t-13}(13)$	$\hat{z}_{t-14}(14)$
95% 概率限 ±				0.26	0.55	0.84	1.15	1.46	1.75	2.04	2.32	2.59	2.84	3.09	3.32	3.58	3.77
50% 概率限 ±				0.09	0.19	0.29	0.39	0.50	0.60	0.70	0.79	0.88	0.97	1.05	1.13	1.22	1.29
	t	z_t	a_t														
	19	23.7															
原点	20	23.4															
	21	23.1	−0.06	23.16													
	22	22.9			22.97												
	23	22.8				22.81											
	24	22.7					22.69										
	25	22.6						22.59									
	26	22.4							22.51								
	27	22.2								22.45							
	28	22.0									22.40						
	29	21.8										22.36					
	30	21.4											22.32				
	31	20.9												22.30			
	32	20.3													22.28		
	33	19.7														22.27	
	34	19.4															22.25
	35	19.3															

一旦新数据 z_{21} 成为可以使用的，我们立刻就产生一组新的预报值并将其填在表中所示预报值的下方。利用 $a_t = z_t - \hat{z}_{t-1}(1)$ 这一事实，每当得到新数据 z_t 时，对应的 a_t 就可计算出 — 用与 a_t 紧邻的左右两个表值相减。也就是一旦 $z_{21} = 32.1$ 是可以使用的，我们就可以插入 a_{21} 对应的表值 -0.06=23.1-23.16.

5.2.2 ψ 权的计算

假设我们需要做提前期为 $1, 2, \cdots, L$ 的预报。为了得到预报值的概率限，并使新的预报值能通过修正旧预报而算出，这样就必须计算权数 $\psi_1, \psi_2, \cdots, \psi_{L-1}$ 。这可以通过利用 (5.1.4) 实现，形式上有

$$\varphi(B)\pi(B) = \theta(B) \tag{5.2.1}$$

在下式中令 B 的同幂次系数相等：

$$\begin{aligned}(1 - \varphi_1 B - \cdots - \varphi_{p+d} B^{p+d})&(1 + \psi_1 B + \psi_2 B^2 + \cdots) \\ &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\end{aligned} \tag{5.2.2}$$

已知 φ 和 θ 的值，通过令 B 的系数相等得到 ψ 为

$$\begin{aligned}\psi_1 &= \varphi_1 - \theta_1 \\ \psi_2 &= \varphi_1 \psi_1 + \varphi_2 - \theta_2 \\ &\vdots \\ \psi_j &= \varphi_1 \psi_{j-1} + \cdots + \varphi_{p+d} \psi_{j-p-d} - \theta_j\end{aligned} \tag{5.2.3}$$

这里， $\psi_0 = 1, \psi_j = 0$ 当 $j < 0$ 及 $\theta_j = 0$ 当 $j > q$ 。若 K 是整数 $p + d - q$ 和 q 中较大者，那么，当 $j > K$ 时 ψ 满足下面差分方程：

$$\psi_j = \varphi_1 \psi_{j-1} + \varphi_2 \psi_{j-2} + \cdot + \varphi_{p+d} \psi_{j-p-d} \tag{5.2.4}$$

因此 ψ 很容易递推计算得出。例如，对适合于序列 C 的模型 $(1 - 1.8B + 0.8B^2)z_t = a_t$ 我们有

$$(1 - 1.8B + 0.8B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

直接令 B^j 的系数相等，或是利用 (5.2.3) 和 (5.2.4)，其中 $\varphi_1 = 1.8$ 和 $\varphi_2 = -0.8$，我们得到

$$\begin{aligned}
&\psi_0 = 1 \\
&\psi_1 = 1.8 \\
&\psi_j = 1.8\psi_{j-1} - 0.8\psi_{j-2} \qquad (j = 2, 3, 4, \cdots)
\end{aligned}$$

因此

$$\begin{aligned}
\psi_2 &= (1.8 \times 1.8) - (0.8 \times 1.0) = 2.44 \\
\psi_3 &= (1.8 \times 2.44) - (0.8 \times 1.8) = 2.95
\end{aligned}$$

等等．本例中的 ψ 值在表 5.1 中的第二行给出．

5.2.3 ψ 权在修正预报值中的应用

我们有兴趣考虑产生预报的另一种方法．利用 (5.1.13) 式，可以把 z_{t+l+1} 在时间原点 $t+1$ 和 t 的预报值 $\hat{z}_{t+1}(l)$ 和 $\hat{z}_t(l+1)$ 分别表示为

$$\begin{aligned}
\hat{z}_{t+1}(l) &= \psi_l a_{t+1} + \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \cdots \\
\hat{z}_t(l+1) &= \qquad\qquad\; + \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \cdots
\end{aligned}$$

两式相减得到

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1} \tag{5.2.5}$$

显然，只需要加上提前 1 期预报误差 a_{t+1} 的某常数倍数，我们就可以把在原点 t 对 z_{t+l+1} 的预报值修正成为在原点 $t+1$ 对 z_{t+l+1} 的预报值，而这个常数乘子就是 ψ_l．

这样就得出一个十分重要的结论．假设我们现在已经有了在原点 t 提前期为 $1, 2, \cdots, L$ 的预报值．一旦有了 z_{t+1}，我们就可以计算 $a_{t+1} = z_{t+1} - \hat{z}_t(l)$，并乘以某一比例，得到在原点 $t+1$ 经修正的预报值即为 $\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1}$，这里 $l = 1, 2, \cdots, L-1$．提前期为 L 的新预报值 $\hat{z}_{t+1}(L)$ 不能用这种方法计算，但是由提前期较少的预报值并利用差分方程可以很容易地得到该预报值．

表 5.2 预报的修正

提前期 $l\to$			1	2	3	4	5	6	7	8	9	10	11	12	13	14
系数 $\psi_l\to$			1.80	2.44	2.95	3.36	3.69	3.95	4.16	4.33	4.46	4.57	4.65	4.72	4.78	4.82
预报特性 $\to$			$\hat{z}_{t-1}(1)$	$\hat{z}_{t-2}(2)$	$\hat{z}_{t-3}(3)$	$\hat{z}_{t-4}(4)$	$\hat{z}_{t-5}(5)$	$\hat{z}_{t-6}(6)$	$\hat{z}_{t-7}(7)$	$\hat{z}_{t-8}(8)$	$\hat{z}_{t-9}(9)$	$\hat{z}_{t-10}(10)$	$\hat{z}_{t-11}(11)$	$\hat{z}_{t-12}(12)$	$\hat{z}_{t-13}(13)$	$\hat{z}_{t-14}(14)$
95% 概率限 ±			0.26	0.55	0.84	1.15	1.46	1.75	2.04	2.32	2.59	2.84	3.09	3.32	3.58	3.77
50% 概率限 ±			0.09	0.19	0.29	0.39	0.50	0.60	0.70	0.79	0.88	0.97	1.05	1.13	1.22	1.29
t	z_t	a_t														
19	23.7															
20	23.4															
21	23.1	−0.06	23.16													
22	22.9		22.86	22.97												
23	22.8			22.67	22.81											
24	22.7				22.51	22.69										
25	22.6					22.39	22.59									
26	22.4						22.29	22.51								
27	22.2							22.21	22.45							
28	22.0								22.15	22.40						
29	21.8									22.10	22.36					
30	21.4										22.06	22.32				
31	20.9											22.03	22.30			
32	20.3												22.00	22.28		
33	19.7													21.99	22.27	
34	19.4														21.96	22.25
35	19.3															21.95

我们再来讨论一下序列 C 的预报. 在表 5.2 中, 时间原点 $t = 21$ 的预报值是在前文表 5.1 中的诸预报值加上修正项而得到. 为了得出这些预报值, 或者可以用直接的方法, 这同在原点 $t=20$ 为得到预报值而进行的计算一样; 或者可以利用修正方程 (5.2.5). ψ 权的值列在表 5.2 中的第二行, 可以很方便地计算.

特别, 当我们一旦知道了 $z_{21}=23.1$, 就可以算出 $a_{21}=23.1-23.16=-0.06$, 于是有

$$\begin{aligned}\hat{z}_{21}(1)&=22.86=22.97+(1.8)(-0.06)\\ \hat{z}_{21}(2)&=22.67=22.81+(2.44)(-0.06)\\ \hat{z}_{21}(3)&=22.51=22.69+(2.95)(-0.06)\end{aligned}$$

等等.

5.2.4 任意提前期预报概率量信限的计算

(5.1.16) 式表明, 在一般情况下, 在任何时间原点 t 的提前 l 期预报误差方差是

$$e_t^2(l)=[z_{t+l}-\hat{z}_t(l)]^2$$

的期望, 即

$$V(l)=(1+\sum_{j=1}^{l-1}\psi_j^2)\sigma_q^2$$

例如, 利用在表 5.1 中的 ψ 权, 序列 C 的函数 $V(l)/\sigma_a^2$ 值在表 5.3 中给出.

假设 a 是正态, 因此当 t 时刻之前的信息已知时, 过程未来值 z_{t+l} 的条件概率分布 $p(z_{t+l}\mid z_t,z_{t-1},\cdots)$ 将是正态, 其均值为 $\hat{z}_t(l)$, 方差为 $(1+\sum_{j=1}^{l-1}\psi_j^2)^{1/2}\sigma_a$. 图 5.1 表示当时间原点 $t=20$ 之前的信息给出时, 序列 C 未来值 z_{21},z_{22},z_{23} 的条件概率分布.

表 5.3 序列 C 的方差函数

l	1	2	3	4	5	6	7	8	9	10
$V(l)/\sigma_a^2$	1.00	4.24	10.19	18.96	30.24	43.86	59.46	76.79	95.52	115.41

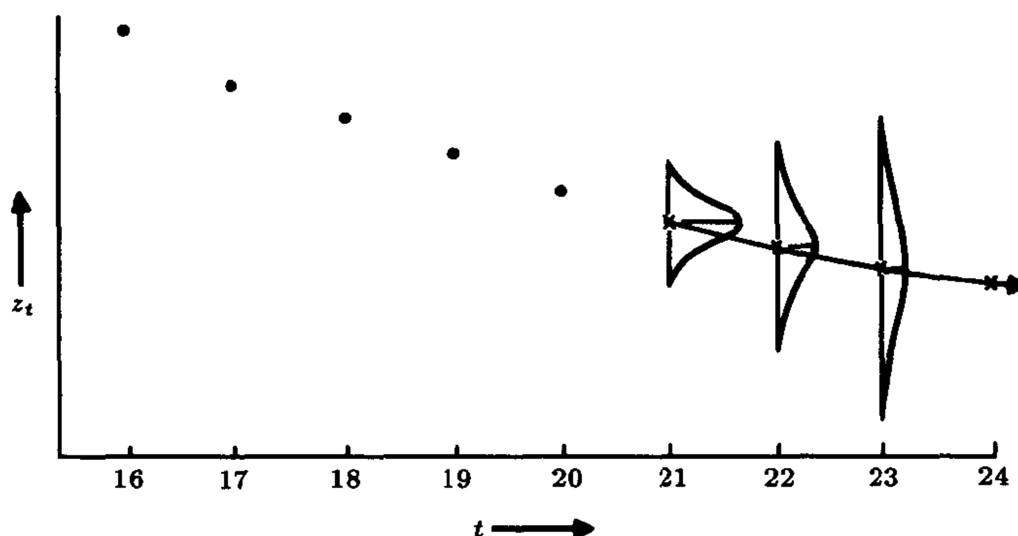


图 5.1　当原点 $t=20$ 之前的信息给出时，序列 C 未来值 z_{21}, z_{22} 和 z_{23} 的条件概率分布

在第 7 章我们将要说明怎样由时间序列数据得到方差 σ_a^2 的估计 s_a. 当观测值的数目不少于 50, 就可以用这种方法求得的 s_a 代替 σ_a, 且 z_{t+l} 的接近于 $1-\epsilon$ 的概率置信限 $z_{t+l}(-)$ 和 $z_{t+l}(+)$ 由下式给出

$$z_{t+l}(\pm) = \hat{z}_t(l) \pm u_{\epsilon/2}\left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} s_a \qquad (5.2.6)$$

其中 $u_{\epsilon/2}$ 是偏差值，标准正态分布以概率 $\epsilon/2$ 超过该值.

在表 7.13 中给出了：对于序列 C,$s_a = 0.134$, 因此，例如对 $\hat{z}_t(2)$ 其 50% 和 95% 的置信限可以给出如下：

50% 置信限：　$\hat{z}_t(2) \pm (0.674)(1 + 1.8^2)^{1/2}(0.134) = \hat{z}_t(2) \pm 0.19$

95% 置信限：　$\hat{z}_t(2) \pm 1.96(1 + 1.8^2)^{1/2}(0.134) = \hat{z}_t(2) \pm 0.55$

在表 5.1 和表 5.2 中，将预报值加上或减去表头的第四、第五行，就得到分别相应于 50% 和 95% 的概率限. 概率限将预报值置于其间.

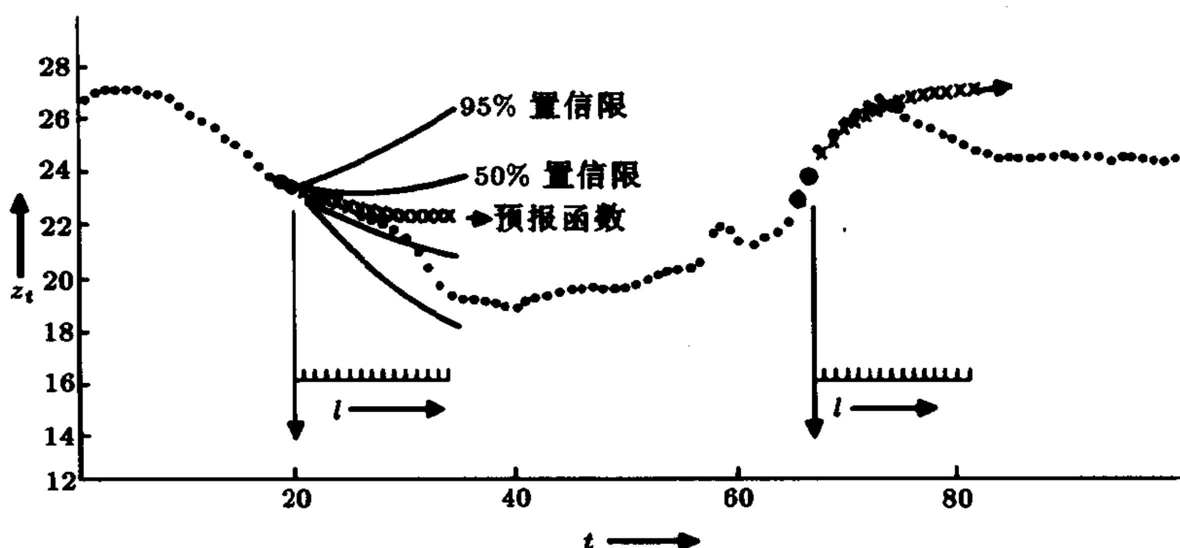


图 5.2 序列 C 的预报和概率限

在图 5.2 中给出序列 C 的一段数据，同时给出以 $t=20$ 和 $t=67$ 为原点的提前若干期预报值 (用 " 十字 " 标明), 还给出了 $z_{20}(l)$ 的 50% 和 90% 的概率限，这里 $l=1,2,\cdots,14$ 。当然我们应该注意到 $z_{t+l}(-)$ 和 $z_{t+l}(+)$ 的解释. 概率限的含义是：在给出了直到原点 t 之前所有信息的条件下，真实值将以 $1-\epsilon$ 的概率出现在上、下置信限之间，即

$$\Pr\{z_{t+l}(-)<z_{t+l}<z_{t+l}(+)\}=1-\epsilon$$

上述解释中的概率只能运用于单个预报值，而不是针对不同提前期的预报值. 例如，对于预报提前期为 10 的概率限，值 z_{t+10} 将以 95% 的概率出现于其间，但是如果希望所有未来值以同样的概率水平出现在同一概率限之间，便是不可能的了。

5.3 预报函数和预报权

直接利用差分方程形式是计算预报值最简便的方法. 单纯从计算的角度来看，其它模式都不太方便. 然而，从研究预报性质的角度出发，更详细地考虑在节 5.1.2 中所讨论的其它的预报形式，特别是考虑预报函数的显式，将是非常有益的.

5.3.1 自回归算子所确定的最终预报函数

在 $t+l$ 时刻 ARIMA 模型可以写成

$$z_{t+l}-\varphi_1 z_{t+l-1}-\cdots-\varphi_{p+d}z_{t+l-p-d}=a_{t+l}-\theta_1 a_{t+l-1}-\cdots-\theta_q a_{t+l-q} \tag{5.3.1}$$

在时刻 t 对 (5.3.1) 式取期望，当 $l>q$ 时我们有

$$\hat{z}_t(l)-\varphi_1\hat{z}_t(l-1)-\cdots-\varphi_{p+d}\hat{z}_t(l-p-d)=0\ (l>q) \tag{5.3.2}$$

其中，对于 $j>0$, 理解为 $\hat{z}_t(-j)=z_{t-j}$. 差分方程 (5.3.2) 的解为

$$\hat{z}_t(l)=b_0^{(t)}f_0(l)+b_1^{(t)}f_1(l)+\cdots+b_{p+d-1}^{(t)}f_{p+d-1}(l) \tag{5.3.3}$$

这里 $l>q-p-d$. 我们注意到预报 $\hat{z}_t(l)$ 就是在第 4 章中引入的余函数。在 (5.3.3) 式中，$f_0(l),f_1(l),\cdots,f_{p+d-1}(l)$ 是预报提前期 l 的函数，通常它们可能包含多项式、指数、正弦和余弦，以及这些函数的乘积。函数 $f_0(l),f_1(l),\cdots,f_{p+d-1}(l)$ 中包含 p 个形如 G^i 和 $D^i\sin(2\pi fl+F)$ 的衰减指数和衰减正弦项，它们分别对应着与平稳自回归算子有关的 $\phi(B)=0$ 的根；这些函数中还包含 d 个多项式项 $l^i,i=0,1,\cdots,d-1$, 多项式为 $d-1$ 次，这是与非平稳差分算子 $\nabla^d=(1-B)^d$ 相联系的。当时间原点 t 给定时，系数 $b_j^{(t)}$ 对于所有的预报提前期 l 都是常数，但是它们将随着原点的改变而变化，使之适应于序列已观测部分的特性。从现在起，我们称 (5.3.3) 式所定义的函数为最终预报函数，所谓“最终”是因为假如恰好是 $q>p+d$, 那么这个函数只能对提前期 $l>q-p-d$ 作预报。

我们从 (5.3.2) 式可以看出，广义自回归算子 $\phi(B)$ 决定了预报函数的数学形式，也就是决定了 (5.3.3) 中 f 的性质。具体说来，即是决定了预报函数是否为多项式、正弦余弦混合形式、指数混合形式，或者为这些函数的组合。

5.3.2 滑动平均算子在确定初值中的作用

自回归算子决定了最终预报函数的性质，而滑动平均算子的作用是确定预报函数，使之"适应"于数据，也就是决定了如何对 $b_0^{(t)}, b_1^{(t)}, \cdots, b_{p+d-1}^{(t)}$ 进行计算和修正.

例如考虑 IMA(0,2,3) 过程

$$z_{t+l} - 2z_{t+l-1} + z_{t+l-2} = a_{t+l} - \theta_1 a_{t+l-1} - \theta_2 a_{t+l-2} - \theta_3 a_{t+l-3}$$

利用节 5.1.2 中关于条件期望的讨论，预报函数定义为

$$\begin{aligned}
\hat{z}_t(1) &= 2z_t - z_{t-1} - \theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2} \\
\hat{z}_t(2) &= 2\hat{z}_t(1) - z_t - \theta_2 a_t - \theta_3 a_{t-1} \\
\hat{z}_t(3) &= 2\hat{z}_t(2) - \hat{z}_t(1) - \theta_3 a_t \\
\hat{z}_t(l) &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \qquad (l > 3)
\end{aligned}$$

由于 $\varphi(B) = (1-B)^2$, 因此最终预报函数是直线

$$\hat{z}(l) = b_0^{(t)} + b_1^{(t)} l \qquad (l > 1)$$

如图 5.3 中所示，该直线通过 $\hat{z}_t(2)$ 和 $\hat{z}_t(3)$ 。然而，如果略去 θ_3 项，则有 $q - p - d = 0$, 于是任意步长的预报值都在过 $\hat{z}_t(1)$ 和 $\hat{z}_t(2)$ 的直线上.

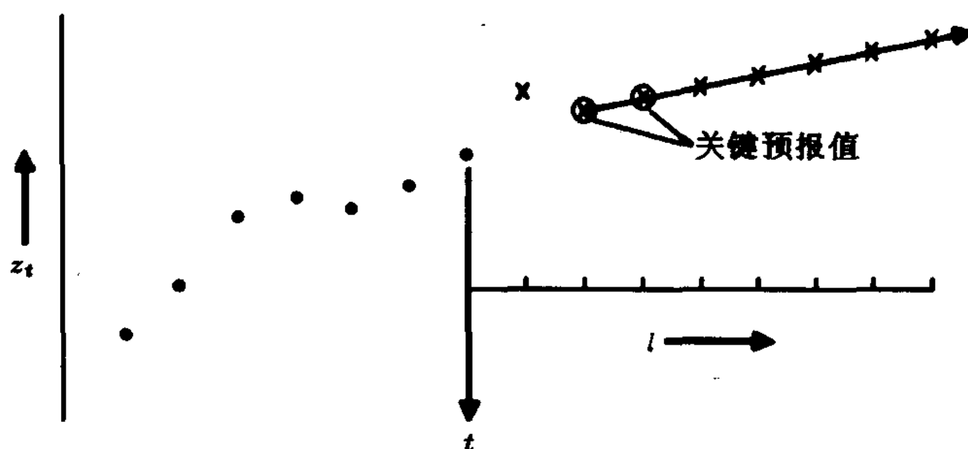


图 5.3 一个 IMA(0,2,3) 过程的最终预报函数

对于一般情形，由于只有一个形如 (5.3.3) 的函数通过 $p+d$ 个点，而最终预报函数就是满足 $\phi(B)$ 所要求形式的唯一曲线，且通过 $p+d$ 个“关键值”$\hat{z}_t(q), \hat{z}_t(q-1), \cdots, \hat{z}_t(q-p-d+1)$，其中 $\hat{z}_t(-j)=z_{t-j}(j=0,1,2,\cdots)$ 。对于 $q=0$ 的极端情形，模型为纯自回归形式 $\varphi(B)z_t=0$，这时最终预报函数曲线通过点 $z_t, z_{t-1}, \cdots$，因此，这些关键值可能包含预报值或序列的真实值，在图中关键值用圆圈标出。

模型中出现的滑动平均”回溯”到观测序列，帮助我们决定预报函数 (由自回归算子 $\varphi(B)$ 所确定) 通过的路径。图 5.4 表示示 (1,1,3) 阶模型 $(1-\phi B)\nabla z_t=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)a_t$ 的一些情况。权函数表明了 $\hat{z}_t(1), \hat{z}_t(2), \hat{z}_t(3)$ 对于 $z_t, z_{t-1}, \cdots$ 的线性依赖关系。由于预报函数包含 $p+d=2$ 个系数，因而它被预报值 $\hat{z}_t(3)$ 和 $\hat{z}_t(2)$(即 $\hat{z}_t(q)$ 和 $\hat{z}_t(q-1)$) 唯一地确定。下面我们将讨论如何确定前文中提到的预报权函数。

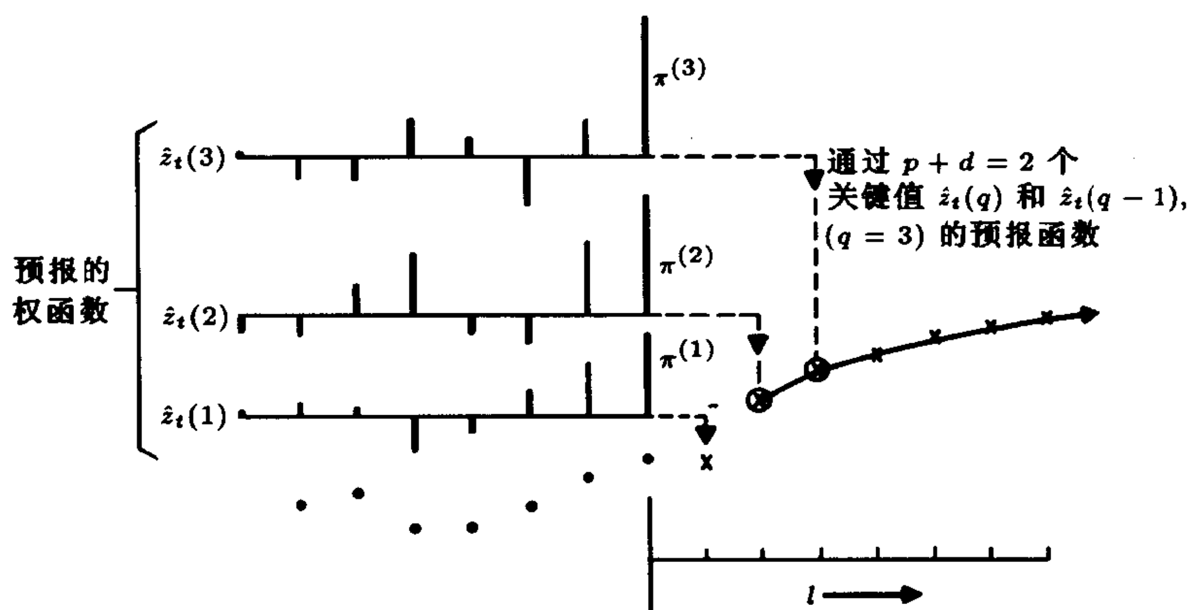


图 5.4 (1,1,3) 过程 $(1-\phi B)\nabla z_t=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)a_t$ 的预报函数对观测值的依赖

5.3.3 提前 l 期预报的权函数

事实上，一般模型也可以写成逆转形式

$$a_t = \pi(B)z_t = (1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)z_t \tag{5.3.4}$$

由此我们可以把预报写为 (5.1.21) 的形式. 在 (5.1.21) 中代之以条件期望，我们得到

$$\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j \hat{z}_t(l-j) \tag{5.3.5}$$

如前所述，对 $h = 0, 1, 2, \cdots, \hat{z}_t(-h) = z_{t-h}$ 。因此，一般有

$$\hat{z}_t(l) = \pi_1 \hat{z}_t(l-1) + \cdots + \pi_{l-1}\hat{z}_t(1) + \pi_l z_t + \pi_{l+1} z_{t-1} + \cdots \tag{5.3.6}$$

特别地，

$$\hat{z}_t(1) = \pi_1 z_t + \pi_2 z_{t-1} + \pi_3 z_{t-2} + \cdots$$

提前多期预报值也可以直接表示成观测值 $z_t, z_{t-1}, z_{t-2}, \cdots$ 的线性函数. 例如，在原点 t 的提前 2 期预报是

$$\begin{aligned}
\hat{z}_t(2) &= \pi_1 \hat{z}_t(1) + \pi_2 z_t + \pi_3 z_{t-1} + \cdots \\
&= \pi_1 \sum_{j=1}^{\infty} \pi_j z_{t-j+1} + \sum_{j=1}^{\infty} \pi_{j+1} z_{t-j+1} \\
&= \sum_{j=1}^{\infty} \pi_j^{(2)} z_{t-j+1}
\end{aligned}$$

其中

$$\pi_j^{(2)} = \pi_1 \pi_j + \pi_{j+1} \qquad (j = 1, 2, \cdots) \tag{5.3.7}$$

依此类推，很容易得出

$$\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} z_{t-j+1} \tag{5.3.8}$$

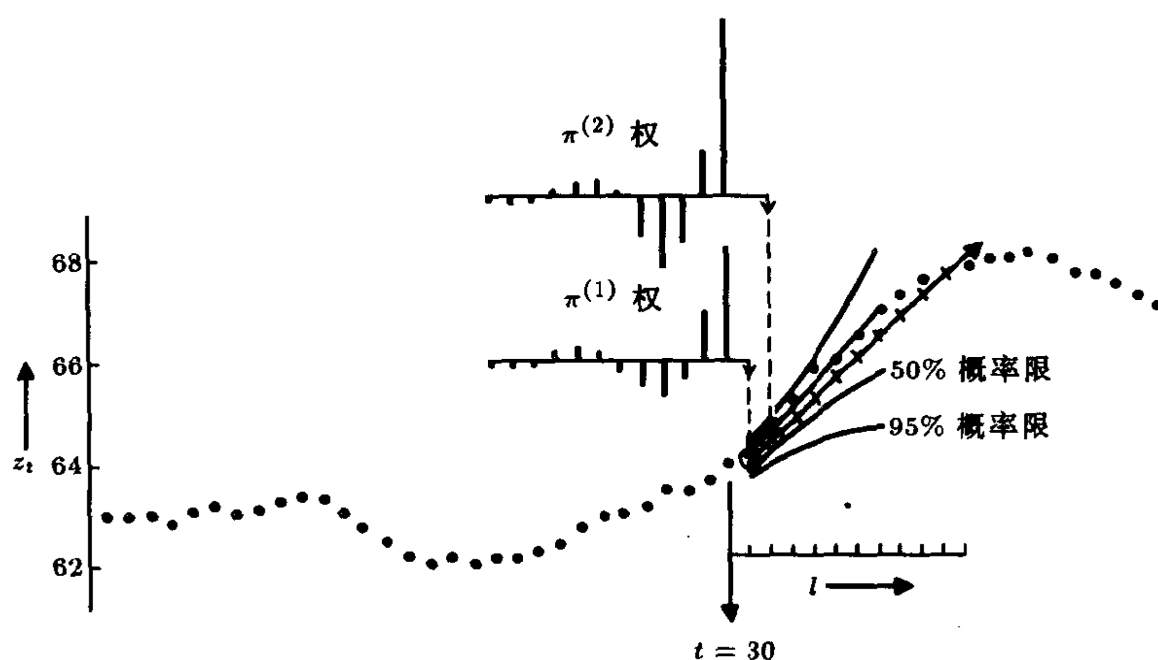


图 5.5 用 $\nabla^2 z_t=(1-0.9B+0.5B^2)a_t$ 拟合某段序列以及在原点 $t=30$ 的预报函数、预报权和概率限

这里

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{h=1}^{l-1} \pi_h \pi_j^{(l-h)} \quad j = 1, 2, \cdots \tag{5.3.9}$$

而 $\pi_j^{(1)} = \pi_j$。计算这些权数的另一种方法在附录 A5.2 中给出。

在 (4.2.22) 和 (5.1.9) 中我们已经看到，权函数 $\pi_1, \pi_2, \pi_3, \cdots$ 本身可以通过比较方程

$$\theta(B)(1 - \pi_1 B - \pi_2 B^2 - \cdots) = \varphi(B)$$

两边的系数得出，有了 π_j 的值，利用 (5.3.9) 或附录 A5.2 的结果就可以很容易地得到 $\pi_j^{(l)}$ 的值.

例如再次考虑模型

$$\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$$

用该模型拟合一个序列，在图 5.5 中给出了其中的一段。在下式

表 5.4 模型 $\nabla^2 z_t(1-0.9B+0.5B^2)a_t$ 的 π 权

j	$\pi_j^{(1)}$	$\pi_j^{(2)}$
1	1.100	1.700
2	0.490	0.043
3	−0.109	−0.463
4	−0.343	−0,632
5	−0.254	−0.336
6	−0.057	0.013
7	0.076	0.181
8	0.097	0.156
9	0.049	0.050
10	−0.004	−0.032
11	−0.028	−0.054
12	−0.023	−0.026

中令系数相等:

$$(1-0.9B+0.5B^2)(1-\pi_1 B-\pi_2 B^2\cdots)=1-2B+B^2$$

得出权 $\pi_j=\pi_j^{(1)}$, 由此并利用 (5.3.7) 式权 $\pi_j^{(2)}$ 也可计算出。在表 5.4 中，对 $j=1,2,\cdots,12$ 给出了这两组权。

在这个例子中，用 $z_t,z_{t-1},\cdots$ 表示提前 1 期和 2 期的预报值分别为

$$\hat{z}_t(1)=1.10z_t+0.49z_{t-1}-0.11z_{t-2}-0.34z_{t-3}-0.25z_{t-4}-\cdots$$
$$\hat{z}_t(2)=1.70z_t+0.43z_{t-1}-0.46z_{t-2}-0.63z_{t-3}-0.34z_{t-4}+\cdots$$

事实上如图 5.5 所示，这些权数遵从衰减的正弦波。

5.4 预报函数及其修正的例子

本节将考虑一般 ARIMA 模型某些特殊情形的预报函数。我们将用在节 5.1.2 中讨论的三种不同形式来表示预报函数。如上所述，由差分方程本身来计算预报是最简单的，其它两种形式对于我们在一些特殊场合了解预报函数的性质是非常有用的。

5.4.1 IMA(0,1,1) 过程的预报

模型为 $\nabla z_t = (1-\theta B)a_t$.

差分方程方法　在 $t+l$ 时刻模型可以写为

$$z_{t+l} = z_{t+l-1} + a_{t+l} - \theta a_{t+l-1}$$

在原点 t 取条件期望，得到

$$\begin{aligned} \hat{z}_t(1) &= z_t - \theta z_t \\ \hat{z}_t(l) &= \hat{z}_t(l-1) \qquad (l > 2) \end{aligned} \tag{5.4.1}$$

因此，在原点 t 的任意提前期预报都在平行于时间轴的直线上，利用 $z_t = \hat{z}_{t-1}(1) + a_t$ 这一事实，我们可以将 (5.4.1) 写成另外两种有用的形式。

第一种形式是

$$\hat{z}_t(l) = \hat{z}_{t-1}(l) + \lambda a_t \tag{5.4.2}$$

这里， $\lambda = 1-\theta$ 。该式表明，先前的预报值 $\hat{z}_{t-1}(l)$ 没有体现出 a_t 值的作用，我们加上 λa_t 对它进行修正。在节 4.3.1 中曾提到，对于任意给定冲击 a_t,λ 用来度量被过程的“水平 ”永远吸收的那部分冲击的比例。因此，把 a_t 的一部分 λa_t(我们预期会被吸收的部分) 增添到预报中是合理的。

第二种改写 (5.4.1) 的方式为

$$\hat{z}_t(l) = \lambda z_t + (1-\lambda)\hat{z}_{t-1}(l) \tag{5.4.3}$$

该式表明，新的预报值是旧预报值和新观测值依变量 λ 的线性内插。由 (5.4.3) 显然可知，若 λ 很小，我们将主要依赖于过去数据的加权平均，而对新观测值 z_t 大打折扣。反之，如果 $\lambda = 1$, 则过去数据完全消匿， $\hat{z}_t(1) = z_t$, 即任何未来时刻的预报就是当前值。当 $\lambda > 1$ 时，在 $\hat{z}_{t-1}(l)$ 和 z_t 之间引入的不是内插而是外推，这时，在 (5.4.2) 中必须增大预报误差以说明预报值的变化。

预报函数的求和形式 最终预报函数是差分方程 $(1-B)\hat{z}_t(l)=0$ 的解，因此 $\hat{z}_t(l)=b_0^{(t)}$，且由于 $q-p-d=0$，可得对于任意步长的预报有

$$\hat{z}_t(l)=b_0^{(t)} \qquad (l>0) \tag{5.4.4}$$

对于任意固定的原点，$b_0^{(t)}$ 为一常数，因此任意提前期的预报都在平行于时间轴的一直线上。然而，系数 $b_0^{(t)}$ 将随着新观测值的获得和原点的推进而加以修正。因此，预报函数可以看作是预报提前期 l 的零阶多项式，其系数适应于时间原点 t。

通过比较 (5.4.4) 和 (5.4.1)，我们看到

$$b_0^{(t)}=\hat{z}_t(l)=z_t-\theta a_t$$

同样地，参照 (4.3.4) 式，由于模型相对于时间原点 k 的截断形式为

$$\begin{aligned}z_t&=\lambda S_{t-k-1}a_{t-1}+a_t+(z_k-\theta a_k)\\&=\lambda(a_t+\cdots+a_{k+1})+a_t+(z_k-\theta a_k)\end{aligned}$$

由此得出

$$z_t(l)=b_0^{(t)}=\lambda S_{t-k}a_t+(z_k-\theta a_k)=\lambda(a_t+\cdots+a_{k+1})+(z_k-\theta a_k)$$

且 $\psi_j=\lambda(j=1,2,\cdots)$，因此，当时间原点从 t 到 $t+1$ 时，自适应系数可按照下式进行修正：

$$b_0^{(t+1)}=b_0^{(t)}+\lambda a_{t+1} \tag{5.4.5}$$

该式类似于 (5.4.2)。

预报值作为既往观测的加权平均 对该过程而言，(5.3.8) 中的 $\pi_j^{(l)}$ 权就是提前 1 期预报的权数，因此，利用 (4.3.6) 我们可写成

$$\hat{z}_t(l)=b_0^{(t)}=\lambda z_t+\lambda(1-\lambda)z_{t-1}+\lambda(1-\lambda)^2z_{t-2}+\cdots \tag{5.4.6}$$

故对于 IMA(0,1,1) 过程，任意提前期的预报值是当前和过去 z 的指数加权滑动平均。

例：序列 A 的预报　在第七章中将说明，模型

$$(1-B)z_t=(1-0.7B)a_t$$

可对序列 A 做精确的拟合。在图 5.6 中给出了在原点 t=39,40,41,42 和 43 以及原点 t=79 提前期为 1,2,$\cdots$,20 的预报。该模型的 π_j 权就是任意提前期预报的权数，在表 5.5 中给出。在图 5.6 中，对于预报值 $\hat{z}_{39}(l)$ 的这些权数用图示的方法给出。

方差函数　对于该模型 $\pi_j=\lambda(j=1,2,\cdots)$，因此，提前 l 期预报方差的表示式 (5.1.16) 成为

$$V(l)=\sigma_a^2[1+(l-1)\lambda^2] \tag{5.4.7}$$

对于序列 A 来说，在 (5.4.7) 中使用估计 $s_a^2=0.101$ 是适宜的，我们算出 50% 和 95% 的概率限并在图 5.6 中对原点 t=79 给出。

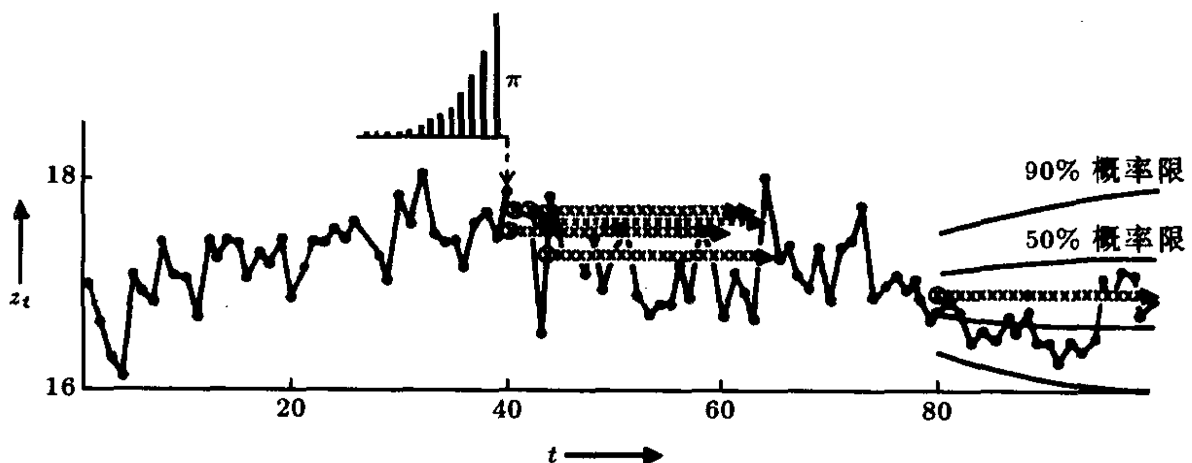


图 5.6 序列 A 的一段在原点 t=39,40,41,42,43 和 t=79 处的预报值

表 5.5 对序列 A 采用模型 $\nabla z_t=(1-0.7B)a_t$ 作任意提前期预报时，对既往 z 使用的权数

j	π_j	j	π_j
1	0.300	7	0.035
2	0.210	8	0.025
3	0.147	9	0.017
4	0.103	10	0.012
5	0.072	11	0.008
6	0.050	12	0.006

5.4.2 IMA(0,2,2) 过程的预报

模型为 $\nabla^2 z_t = (1-\theta_1 B - \theta_2 B^2)a_t$.

差分方程方法　在 $t+l$ 时刻模型可以写成

$$z_{t+l} = 2z_{t+l-1} - z_{t+l-2} + a_{t+l} - \theta_1 a_{t+l-1} - \theta_2 a_{t+l-2}$$

在时刻 t 取条件期望，我们得到

$$\begin{aligned}\hat{z}_t(1) &= 2z_t - z_{t-1} - \theta_1 a_t - \theta_2 a_{t-1}\\ \hat{z}_t(2) &= 2\hat{z}_t(1) - z_t - \theta_2 a_t\\ \hat{z}_t(l) &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \qquad (l > 3)\end{aligned}$$

由此便可以计算出预报值. 图 5.5 中的序列就是用这种方式所做的预报，在节 5.1.2 中我们已加以说明. 产生 L 个预报值中前 $L-1$ 个的另一种方法是通过公式 (5.2.5) 加以修正:

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1} \tag{5.4.8}$$

如 (4.3.15) 式一样，截断求和模型是

$$\begin{aligned}z_t = \lambda_0 S_{t-k-1} a_{t-1} + \lambda_1 S^{(2)}_{t-k-1} a_{t-1} + a_t + b_0^{(k)} + b_1^{(k)}(t-k)\\ \lambda_0 = 1+\theta_2, \lambda_1 = 1-\theta_1-\theta_2\end{aligned} \tag{5.4.9}$$

故 $\psi_j = \lambda_0 + j\lambda_1 (j=1,2,\cdots)$, 因此该模型的修正函数为

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + (\lambda_0 + l\lambda_1)a_{t+l} \tag{5.4.10}$$

求和形式的预报　最终预报函数是 $(1-B)^2\hat{z}_t(l)=0$ 的解，因此 $\hat{z}_t(l)=b_0^{(t)}+b_1^{(t)}l$。由于 $q-p-d=0$，最终预报函数给出的任意提前期预报为

$$\hat{z}_t(l)=b_0^{(t)}+b_1^{(t)}l \qquad (l>0) \tag{5.4.11}$$

因此，预报函数是预报提前期 l 的线性函数，相应系数是依赖于时间原点 t 的。随机模型的截断求和形式为

$$z_{t+l}=\lambda_0 S_{t+l-k-1}a_{t+l-1}+\lambda_1 S_{t+l-k-1}^{(2)}+a_{t+l}+b_0^{(k)}+b_1^{(k)}(t+l-k)$$

在原点 t 取条件期望，我们得到

$$\begin{aligned}\hat{z}_t(l)&=\lambda_0 S_{t-k}a_t+\lambda_1(la_t+(l+1)a_{t-1}\\&\quad+\cdots+(l+t-k-1)a_{k+1})+b_0^{(k)}+b_1^{(k)}(t+l-k)\\&=[\lambda_0 S_{t-k}a_t+\lambda_1 S_{t-k-1}^{(2)}a_{t-1}+b_0^{(k)}+b_1^{(k)}(t-k)]\\&\quad+(\lambda_1 S_{t-k}a_t+b_1^{(k)})l\end{aligned}$$

修正系数可定义为

$$\begin{aligned}b_0^{(t)}&=\lambda_0 S_{t-k}a_t+\lambda_1 S_{t-k-1}^{(2)}+b_0^{(k)}+b_1^{(k)}(t-k)\\b_1^{(t)}&=\lambda_1 S_{t-k}a_t+b_1^{(k)}\end{aligned} \tag{5.4.12}$$

或形式上将无限求和形式记为 $b_0^{(t)}=\lambda_0 Sa_t+\lambda_1 S^2 a_{t-1}$ 和 $b_1^{(t)}=\lambda_1 Sa_t$，由此得到其修正公式是

$$\begin{aligned}b_0^{(t)}&=b_0^{(t-1)}+b_1^{(t-1)}+\lambda_0 a_t\\b_1^{(t)}&=b_1^{(t-1)}+\lambda_1 a_t\end{aligned} \tag{5.4.13}$$

该式与关系式 (4.3.17) 类似. 在 $b_0^{(t)}$ 的修正公式中增添了一项 $b_1^{(t-1)}$，通过斜率的调整以改变位置参数 b_0 的值，使之适合于新的时间原点。还应该注意到， λ_0 和 λ_1 分别是冲击 a_t 被传递进入位置和斜率参数的系数。

预报值作为既往观测的加权平均　对于该模型而言，预报函数是通过预报值 $\hat{z}_t(1)$ 和 $\hat{z}_t(2)$ 的直线。在图 5.5 中对这类序列的预报加以说明，预报值是在原点 $t = 30$ 作出，并具有适当的权函数。可以看出，整个预报函数对预报值 $\hat{z}_t(1)$ 和 $\hat{z}_t(2)$ 的依赖关系，恰反映了预报值对于序列既往值 z 的依赖关系。$\hat{z}_t(1)$ 和 $\hat{z}_t(2)$ 的权函数在图中表示出，并在表 5.4 中给出了相应的值。

这个例子再次说明了 AR 算子 $\varphi(B)$ 决定了预报所采用函数的形式 (如在本例中是一直线), 而 MA 算子对于决定拟合既往数据的函数形式起着重要作用。

预报函数的自适应系数对既往 z 的依赖关系　对于一般模型来说, 由于预报函数自适应系数的值是由 $\hat{z}_t(q), \hat{z}_t(q-1), \cdots, \hat{z}_t(q-p-d+1)$ 所决定的，而这些值又能表示为观测值的函数，因此，自适应系数本身也可以表示为观测的函数。

例如对图 5.5 中的模型 $\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$, 有

$$\hat{z}_t(1) = b_0^{(t)} + b_1^{(t)} = \sum_{j=1}^{\infty} \pi_j^{(1)} z_{t-j+1}$$

$$z_t(2) = b_0^{(t)} + 2b_1^{(t)} = \sum_{j=1}^{\infty} \pi_j^{(2)} z_{t-j+1}$$

因此

$$b_0^{(t)} = 2\hat{z}_t(1) - \hat{z}_t(2) = \sum_{j=1}^{\infty} (2\pi_j^{(1)} - \pi_j^{(2)}) z_{t-j+1}$$

及

$$b_1^{(t)} = \hat{z}_t(2) - \hat{z}_t(1) = \sum_{j=1}^{\infty} (\pi_j^{(2)} - \pi_j^{(1)}) z_{t-j+1}$$

图 5.7 中给出了权函数的图形。

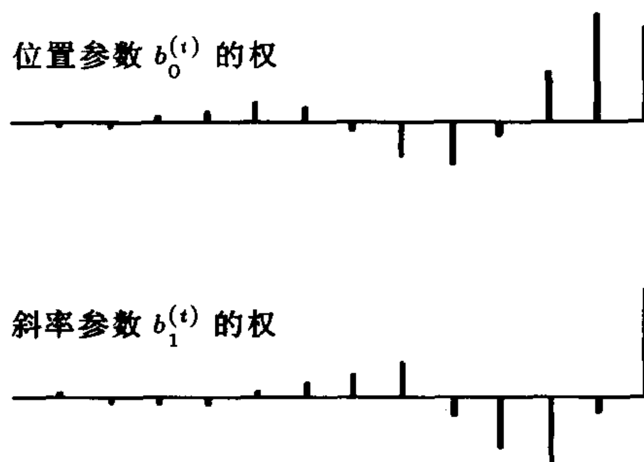


图 5.7 模型 $\nabla^2 z_t=(1-0.9B+0.5B^2)a_t$, 为确定位置和斜率对既往值使用的权数

预报误差的方差　利用 (5.1.16) 及 $\psi_j = \lambda_0 + j\lambda_1$, 提前 l 期预报误差的方差是

$$V(l) = \sigma_a^2[1+(l-1)\lambda_0^2+\frac{1}{6}l(l-1)(2l-1)\lambda_1^2+\lambda_0\lambda_1 l(l-1)] \quad (5.4.14)$$

利用估计值 $s_a^2 = 0.032, \lambda_0 = 0.5$, 及 $\lambda_1 = 0.6$, 在图 5.5 中给出了在原点 $t = 30$ 各期预报的 50% 和 95% 的概率限。

5.4.3 一般 IMA$(0, d, q)$ 过程的预报

作为例子，考虑 (0,1,3) 阶过程:

$$(1 - B)z_{t+1} = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_{t+1}$$

在时刻 t 取条件期望，我们得到

$$\begin{aligned}
\hat{z}_t(1) - z_t &= -\theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2} \\
\hat{z}_t(2) - \hat{z}_t(1) &= -\theta_2 a_t - \theta_3 a_{t-1} \\
\hat{z}_t(3) - \hat{z}_t(2) &= -\theta_3 a_t \\
\hat{z}_t(l) - \hat{z}_t(l-1) &= 0 \qquad (l = 4, 5, 6, \cdots)
\end{aligned}$$

由于 $q - p - d = 2$, 不出所料，对一切 $l > 2$ 有 $\hat{z}_t(l) = \hat{z}_t(3) = b_0^{(t)}$. 如图 5.8 中所示，在达到最终预报函数的水平之前，预报函数有两次初始的“跳跃”, 跳跃值依赖于既往的 a 。

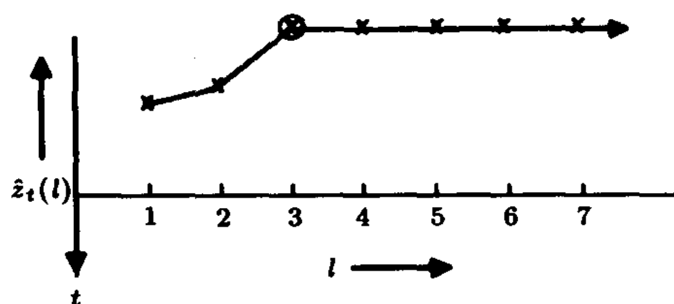


图 5.8 一个 IMA(0,1,3) 过程的预报函数

对于 IMA$(0,d,q)$ 过程，最终预报函数满足 $(1-B)^d\hat{z}_t(l)=0$，其解为 l 的 $d-1$ 阶多项式：

$$\hat{z}_t(l)=b_0^{(t)}+b_1^{(t)}l+b_2^{(t)}l^2+\cdots+b_{d-1}^{(t)}l^{d-1}$$

该式给出了 $l>q-d$ 时的预报值 $\hat{z}_t(l)$．随着原点改变，系数 $b_0^{(t)},b_1^{(t)},\cdots,b_{d-1}^{(t)}$ 必须不断地加以修正。原点 t 的预报值将会做 $q-d$ 次"跳跃"，跳跃值依赖于 $a_t,a_{t-1},\cdots a_{t-q+1}$，此后，预报值将遵从前述多项式。

5.4.4 自回归过程的预报

考虑 $(p,d,0)$ 阶过程

$$\varphi(B)z_t=a_t$$

最终预报函数是 $\varphi(B)\hat{z}_t(l)=0$ 的解。它适用于任意提前期的预报且通过序列的最后 $p+d$ 个有效值。例如，IBM 股票的模型 (序列 B) 十分接近于

$$(1-B)z_t=a_t$$

因此

$$\hat{z}_t(l)\simeq z_t$$

这表明，对于任何未来时刻的最好的预报值非常接近于该股票的当前值。$\hat{z}_t(l)$ 的权函数是在时刻 t 的尖峰，而不是历史值的平均。

平稳自回归模型　考虑 $(p,0,0)$ 阶的过程 $\phi(B)\tilde{z}_t = a_t$, 这里，$\phi(B)$ 是平稳算子，$\tilde{z}_t = z_t - \mu$, 而 $E[z_t] = \mu$。一般说来，该过程所产生的预报函数是指数和衰减正弦的混合形式。特别对于 $p = 1$, 即 (1,0,0) 阶模型：

$$(1 - \phi B)\tilde{z}_t = a_t \qquad (-1 < \phi < 1)$$

对于所有 $l > 0$, 其预报函数是 $(1 - \phi B)\hat{\tilde{z}}_t(l) = 0$ 的解，即

$$\hat{\tilde{z}}_t(l) = b_0^{(t)}\phi^l \qquad (l > 0) \tag{5.4.15}$$

也就是 $\hat{\tilde{z}}_t(l) = \phi\tilde{z}_t$, 因此 $b_0^{(t)} = \tilde{z}_t$ 及

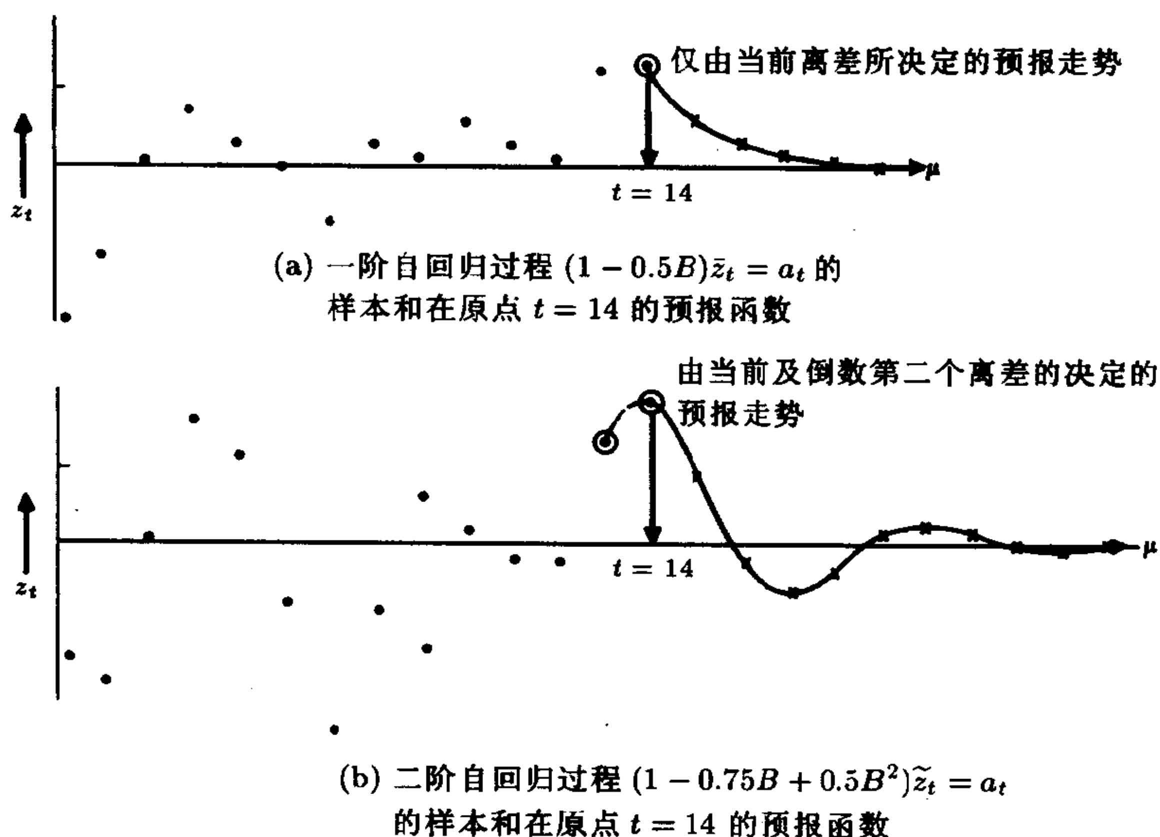
$$\hat{\tilde{z}}_t(l) = \tilde{z}_t\phi^l$$

故对于原过程 z_t 的预报值是 $\hat{z}_t(l) = \mu + \phi^l(z_t - \mu)$。

因此最小均方误差预报表明预报预示着当前离差将从均值按指数衰减到零。图 5.9(a) 给出了由过程 $(1 - 0.5B)\tilde{z}_t = a_t$ 产生的时间序列及在原点 $t = 14$ 的预报函数。该函数的走势看来是由离差 $\tilde{z}_{14}$ 完全确定。类似地，对于二阶自回归过程的最小均方误差预报也是如此，当前离差从均值衰减到零，或是按照衰减正弦波或是按照两个指数衰减的混合形式。图 5.9(b) 给出了过程 $(1 - 0.75B + 0.5B^2)\tilde{z}_t = 0$ 产生的时间序列及在原点 $t = 14$ 的预报值。因此，预报函数在原点 t 的走势由最后两个离差 $\tilde{z}_{14}$ 和 $\tilde{z}_{15}$ 完全确定。

(1,0,0) 过程预报的方差函数　为了进一步说明 (5.1.16) 的应用，我们来推导一阶自回归过程的方差函数。由于在 $t+l$ 时刻模型可以写成

$$\tilde{z}_{t+l} = a_{t+l} + \phi a_{t+l-1} + \cdots + \phi^{l-1}a_{t+l} + \phi^l\tilde{z}_t$$



(a) 一阶自回归过程 $(1-0.5B)\tilde{z}_t=a_t$ 的样本和在原点 $t=14$ 的预报函数

(b) 二阶自回归过程 $(1-0.75B+0.5B^2)\widetilde{z}_t=a_t$ 的样本和在原点 $t=14$ 的预报函数

图 5.9　一阶和二阶自回归过程的预报函数

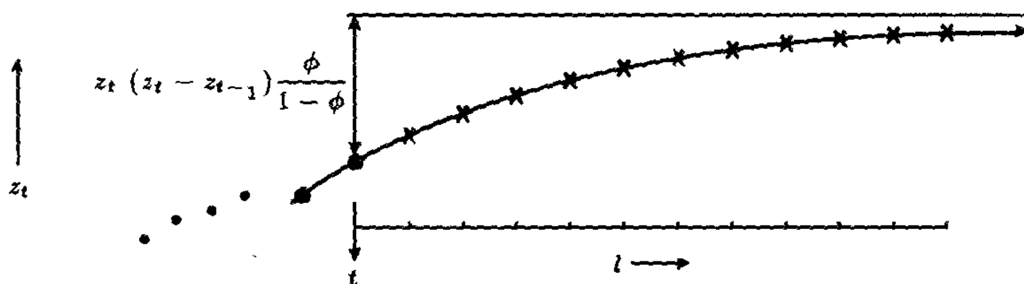
由 (5.4.15) 得到

$$e_t(l)=\widetilde{z}_{t+l}-\hat{\widetilde{z}}_t(l)=a_{t+l}+\phi a_{t+l-1}+\cdots+\phi^{l-1}a_{t+1}$$

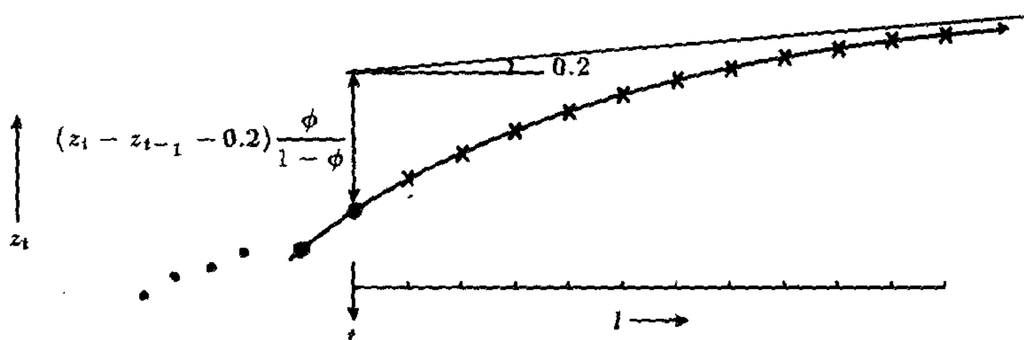
因此

$$\begin{aligned}V(l)=\operatorname{var}[e_t(l)]&=\sigma_a^2(1+\phi^2+\cdots+\phi^{2l-2})\\&=\frac{\sigma_a^2(1-\phi^{2l})}{1-\phi^2}\end{aligned}\tag{5.4.16}$$

我们看到，对于该平稳过程而言，当 l 趋于无穷时，方差增加到常数 $\gamma_0=\sigma_1^2/(1-\phi^2)$，这与过程的方差在最终预报值附近变化相



(a) (1, 1, 0) 过程 $(1-0.8B)\nabla z_t = a_t$ 的预报



(b) (1, 1, 0) 过程 $(1-0.8B)(\nabla z_t - 0.2) = a_t$ 的预报

图 5.10 (1,1,0) 过程的预报函数

符。非平稳模型预报方差函数的特性与此相反，当预报提前期较多时预报方差函数会突然“爆炸”。

$(p, d, 0)$ 阶非平稳自回归模型　考虑模型

$$\phi(B)\nabla^d z_t = a_t$$

对于该过程作 d 次差分，前几步预报呈上升，随后便回落到均值。若不加声明，通常假定 $\nabla^d z_t$ 的均值为零。正象我们在第 4 章中所讨论的，如果需要的话可以在模型中引入非零均值，用 $\nabla^d z_t$ 代替 $\nabla^d z_t - \mu_w$。作为例子，考虑模型

$$(1-\phi B)(\nabla z_t - \mu_w) = a_t \tag{5.4.17}$$

用 $t+j$ 代替 t，并在原点 t 取条件期望，很容易得到 (与 (5.4.15) 式对比)

$$\hat{z}_t(j) - \hat{z}_t(j-1) - \mu_w = \phi^j(z_t - z_{t-1} - \mu_w)$$

或 $\hat{w}_t(j)-\mu_w=\phi^j(w_t-\mu_w)$, 其中 $w_t=\nabla z_t$, 该式表明被预报的差分如何从初值 $w_t=z_t-z_{t-1}$ 依指数衰减到其均值 μ_w 。对于该式从 $j=1$ 到 $j=l$ 求和，即利用 $\hat{z}_t(l)=\hat{w}_t(l)+\cdots+\hat{w}_t(1)+z_t$, 我们得到预报函数为

$$\hat{z}_t(l)=z_t+\mu_w l+(z_t-z_{t-1}-\mu_w)\frac{\phi(1-\phi^l)}{1-\phi} \qquad (l \geqslant 1)$$

上式渐近地逼近直线

$$f(l)=z_t+\mu_w l+(z_t-z_{t-1}-\mu_w)\frac{\phi}{1-\phi}$$

其中决定性的斜率为 μ_w 。图 5.10 给出了 $\phi=0.5,\mu_w=0$ 和 $\phi=0.8,\mu_2=0.2$ 两种情形的预报值。我们在第 7 章中将说明，$\phi=0.8,\mu_w=0$ 的模型 (5.4.17) 可以近似地表示序列 C, 而在图 5.1 和图 5.2 中我们已经说明了有关模型的预报。现在我们来考虑一些重要混合模型的预报。

5.4.5 (1,0,1) 过程的预报

差分方程方法　考虑平稳模型

$$(1-\phi B)\tilde{z}_t=(1-\theta B)a_t$$

其预报很容易由下式得出

$$\begin{aligned}\hat{\tilde{z}}_t(1)&=\phi\tilde{z}_t-\theta a_t\\ \hat{\tilde{z}}_t&=\phi\tilde{z}_t(l-1) \qquad (l \geqslant 2)\end{aligned} \tag{5.4.18}$$

上面预报值按照几何级数衰减到均值，这与一阶自回归过程相同，但是它的提前 1 期预报用一个依赖于 $a_t=z_t-\hat{z}_{t-1}(1)$ 的因子来改进。ψ 权为

$$\psi_j=(\phi-\theta)\phi^{j-1} \qquad (j=1,2,\cdots)$$

因此，利用 (5.2.5) 式，提前期为 $1,2,\cdots,L-1$ 的修正预报可以由提前期为 $2,3,\cdots,L$ 的先前预报值得到，即按照下式

$$\hat{\hat{z}}_{t+1}(l)=\hat{\hat{z}}_t(l+1)+(\phi-\theta)\phi^{l-1}a_{t+1}$$

求和形式　对于一切 $l>0$, 最终预报函数是 $(1-\phi B)\hat{\hat{z}}_t(l)=0$ 的解，即

$$\hat{\hat{z}}_t(l)=b_0^{(t)}\phi^l \qquad (l>0)$$

从而

$$\hat{\hat{z}}_t(1)=b_0^{(t)}\phi=\phi\tilde{z}_t-\theta a_t=\left[(1-\frac{\theta}{\phi})\tilde{z}_t+\frac{\theta}{\phi}\hat{\hat{z}}_{t-1}(1)\right]\phi$$

因此

$$\hat{\hat{z}}_t(l)=\left[(1-\frac{\theta}{\phi})\tilde{z}_t+\frac{\theta}{\phi}\hat{\hat{z}}_{t-1}(1)\right]\phi^l \tag{5.4.19}$$

由此可知，提前 l 期预报离差从初始值起按指数衰减，而预报离差是先前的 1 期预报离差与当前离差之间的线性插值。当 ϕ 等于 1 时，任意提前期的预报都成为熟知的指数加权滑动平均，而 (5.4.19) 式就等于 (5.4.3) 式。

用于既往观测值的权　π 权 —— 即是为获得提前 1 期预报时对既往观测值赋予的权 —— 是

$$\pi_j=(\phi-\theta)\theta^{j-1} \qquad (j=1,2,\cdots)$$

注意，对于该平稳过程来说权数之和等于 $(\phi-\theta)(1-\theta)$ 而不等于 1. 如果 ϕ 等于 1, 该过程成为非平稳的 IMA(0,1,1) 过程，则权数之和等于 1, 且由该过程产生的序列的性状与 z_t 的水平是独立的。

例如，稍后我们将用 $\phi=0.9$ 和 $\theta=0.6$ 的 (1,0,1) 模型拟合序列 A, 所得权数为 $\pi_1=0.30,\pi_2=0.18,\pi_3=0.11,\pi_4=0.07,\cdots,$

权数之和为 0.75. 预报值 (5.4.19) 缓慢地衰减到均值，当预报步长较短时，其预报值与由另一 IMA(0,1,1) 模型 $\nabla z_t = a_t - 0.7a_{t-1}$ 得到的预报值难以区分，后者的权数是 $\pi_1 = 0.30, \pi_2 = 0.21, \pi_3 = 0.15, \pi_4 = 0.10$ 等等，其和等于 1. 后一模型的优点在于它不是把模型束缚于固定的均值。

方差函数　由于 ψ 权由下式给出

$$\psi_j = (\phi - \theta)\phi^{j-1} \qquad (j = 1, 2, \cdots)$$

可得到方差函数为

$$V(l) = \sigma_a^2 \left[1 + (\phi - \theta)^2 \frac{1 - \phi^{2l-2}}{1 - \phi^2} \right] \tag{5.4.20}$$

该值渐近增加到 $\sigma_a^2(1 - 2\phi\theta + \theta^2)/(1 - \phi^2)$, 这正是过程的方差 γ_0 .

5.4.6 (1,1,1) 过程的预报

另一种重要的混合模型就是非平稳的 (1,1,1) 过程

$$(1 - \phi B)\nabla z_t = (1 - \theta B)a_t$$

差分方程方法　在 $t + l$ 时刻模型可以写成

$$z_{t+l} = (1 + \phi)z_{t+l-1} - \phi z_{t+l-2} + a_{t+l} - \theta a_{t+l-1}$$

取条件期望，我们得到

$$\begin{aligned} \hat{z}_t(1) &= (1 + \phi)z_t - \phi z_{t-1} - \theta a_t \\ \hat{z}_t(l) &= (1 + \phi)\hat{z}_t(l - 1) - \phi \hat{z}_t(l - 2) \qquad (l > 1) \end{aligned} \tag{5.4.21}$$

求和形式　由于 $q < p + d$, 对一切 $l > 0$ 最终预报函数是 $(1 - \phi B)(1 - B)\hat{z}_t(l) = 0$ 的解，即

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)}\phi^l$$

将 $\hat{z}_t(1)$ 和 $\hat{z}_t(2)$ 代入 (5.4.21), 显然可得

$$b_0^{(t)} = z_t + \frac{\phi}{1-\phi}(z_t - z_{t-1}) - \frac{\theta}{1-\phi}a_t$$
$$b_1^{(t)} = \frac{\theta a_t - \phi(z_t - z_{t-1})}{1-\phi}$$

因此最终得到

$$\hat{z}_t(l) = z_t + \phi\frac{1-\phi^l}{1-\phi}(z_t - z_{t-1}) - \theta\frac{1-\phi^l}{1-\phi}a_t \tag{5.4.22}$$

对于较大的 l, 显然预报值趋于 $b_0^{(t)}$.

对既往观测值加权　在 (5.4.22) 中消去 a_t, 我们得到用既往观测值表示的另一种预报形式:

$$\hat{z}_t(l) = \left[1 - \frac{\theta-\phi}{1-\phi}(1-\phi^l)\right] z_t + \left[\frac{\theta-\phi}{1-\phi}(1-\phi^l)\right] \bar{z}_{t-1}(\theta) \tag{5.4.23}$$

这里, $\bar{z}_{t-1}(\theta)$ 是关于参数 θ 的指数加权滑动平均, 即 $\bar{z}_{t-1}(\theta) = (1-\theta)\sum_{j=1}^{\infty}\theta^{j-1}z_{t-j}$ 。因此, 该过程的 π 权只在 t 时刻有一单峰, 从 $t-1$ 时刻开始就是指数加权滑动平均 (EWMA) 。如果我们称 $(1-\alpha)x+\alpha y$ 为 x 和 y 之间以 α 为变量的线性插值, 那么, (5.4.23) 就是 z_t 和 $\bar{z}_{t-1}(\theta)$ 之间的线性插值. 当提前期为 1 时, 插值变量是 $\theta-\phi$, 而当预报提前期增大时, 该变量趋于 $(\theta-\phi)/(1-\phi)$ 。例如, 当 $\theta=0.9$ 和 $\phi=0.5$ 时, 提前 1 期预报是

$$\hat{z}_t(1) = 0.6z_t + 0.4\bar{z}_{t-1}(\theta)$$

预报提前期加大时, 预报值接近于

$$\hat{z}_t(\infty) = 0.2z_t + 0.8\bar{z}_{t-1}(\theta)$$

5.5 状态空间模型公式用于精确预报

5.5.1 ARIMA 过程的状态空间模型表示

为了预报，同时也为了模型的设定和参数的极大似然估计，用状态空间形式来表示 ARIMA 模型近来引起了极大定兴趣. 如在节 5.1 中一样，对于 ARIMA(p,q) 过程 $\phi(B)z_t=\theta(B)a_t$, 我们定义预报值为 $\hat{z}_t(j)=E_t[z_{t+j}]$, $j=0,1,\cdots,r$, 其中 $r=\max(p+d,q+1)$, 及 $\hat{z}_t(0)=z_t$ 。

由修正方程 (5.2.5) 我们有 $\hat{z}_t(j-1)=\hat{z}_{t-1}(j)+\psi_{j-1}a_t, j=1,2,\cdots,r-1$ 。此外，对于 $j=r>q$ ，联系到 (5.3.2) 我们有 $\hat{z}_t(j-1)=\hat{z}_{t-1}(j)+\psi_{j-1}a_t=\sum\limits_{i=1}^{p+d}\varphi_i\hat{z}_{t-l}(j-i)+\psi_{j-1}a_t$ 。因此，我们将 t 时刻的"状态"向量 $\boldsymbol{Y}_t$ 定义为具有 r 个分量： $\boldsymbol{Y}_t=(z_t,\hat{z}_t(1),\cdots,\hat{z}_t(r-1))'$ 。由以上关系我们看出 $\boldsymbol{Y}_t$ 满足方程

$$\boldsymbol{Y}_t=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & \cdots & 1\\ \varphi_r & \varphi_{r-1} & \cdots & \cdots & \varphi_1\end{bmatrix}\boldsymbol{Y}_{t-1}+\begin{bmatrix}1\\ \psi_1\\ \vdots\\ \vdots\\ \psi_{r-1}\end{bmatrix}a_t$$

这里，当 $i>p+d$ 时， $\varphi_i=0$ 。我们有

$$\boldsymbol{Y}_t=\boldsymbol{\Phi}\boldsymbol{Y}_{t-1}+\boldsymbol{\Psi}a_t \tag{5.5.1}$$

同时还有量测方程

$$Z_t=z_t+N_t=[1,0,\cdots,0]\boldsymbol{Y}_t+N_t=\boldsymbol{H}\boldsymbol{Y}_t+N_t \tag{5.5.2}$$

假定过程 z_t 是在附加白噪声的条件下而被观测到, 则上式中的附加噪声 N_t 存在，否则便认为 $z_t=\boldsymbol{H}\boldsymbol{Y}_t$ 。上述两个方程就构成了所谓的模型状态空间表示，它由状态方程 (或称传递方程)(5.5.1)

和量测方程 (5.5.2) 组成，而 $\boldsymbol{Y}_t$ 就是所谓的状态向量。我们注意到，存在许多不同结构的状态向量，由此也能得出具有一般形式 (5.5.1)—(5.5.2) 的状态空间模型，也就是说 ARIMA 模型的状态空间形式不是唯一的。上述形式的两个方程一般描述所谓的状态空间模型，其中含有非观测的状态向量 $\boldsymbol{Y}_t$ 和观测值 z_t。这两个方程也会在比 ARIMA 模型范围更广的时间序列背景中出现。

我们考虑状态空间模型稍许一般的形式，其状态方程为

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}_t \boldsymbol{Y}_{t-1} + \boldsymbol{a}_t \tag{5.5.3}$$

量测方程为

$$Z_t = \boldsymbol{H}_t \boldsymbol{Y}_t + N_t \tag{5.5.4}$$

其中 $\boldsymbol{a}_t$ 和 N_t 是相互独立的白噪声过程，$\boldsymbol{a}_t$ 是协方差阵为 Σ_a 的向量白噪声过程，而 N_t 是方差为 σ_N^2 的白噪声。在此模型中状态向量 (非观测的)$\boldsymbol{Y}_t$ 概括了动态系统在时刻 t 的状态，状态方程刻画了动态系统随时间的演变，而量测方程 (5.5.4) 表明，观测值 Z_t 是由受附加白噪声干扰的状态变量的线性组合所生成。在 (5.5.3) 中的矩阵 $\boldsymbol{\Phi}_t$ 是 $r \times r$ 的状态转移矩阵，(5.5.4) 中的 $\boldsymbol{H}_t$ 是 $1 \times r$ 的向量，它们都可随时间而变化。通常在应用中都是作为常数矩阵，$\boldsymbol{\Phi}_t = \boldsymbol{\Phi}$ 及 $\boldsymbol{H}_t = \boldsymbol{H}$, 即对于所有 $t > 0$, 这些矩阵都不依赖于 t,ARIMA 模型的状态空间形式 (5.5.1)—(5.5.2) 便是如此。在这种情形，系统或模型称为非时变的。在状态空间模型中状态向量 $\boldsymbol{Y}_t$ 的最小维数必须足够大，以使系统的动态特征能用形如 (5.5.3) 的简单的一阶 Markov 结构描述。

5.5.2 用于预报的 Kalman 滤波关系式

对于状态空间模型 (5.5.3)—(5.5.4) 来说，就是在有限过去时段观测值 $Z_t, \cdots, Z_1$ 的基础上，确定状态向量 $\boldsymbol{Y}_t$ 的有限样本最优 (均方误差阵为最小) 估计，即确定出 $\hat{\boldsymbol{Y}}_{t+l|t} = E[\boldsymbol{Y}_{t+l} \mid Z_t, \cdots, Z_l]$, , 并以 $\boldsymbol{V}_{t+l|t} = E[(\boldsymbol{Y}_{t+l} - \hat{\boldsymbol{Y}}_{t+l|t})(\boldsymbol{Y}_t + l - \hat{\boldsymbol{Y}}_{t+l|t})']$ 为其误差协方差阵。所谓 Kalman 滤波方程是一种方便的计算方法，特别对于

获得当前估计值 $\hat{\boldsymbol{Y}}_{t|t}$ 也是有效的。如所周知，从某些适当的初值 $\boldsymbol{Y}_0 = \hat{\boldsymbol{Y}}_{0|0}$ 及 $\boldsymbol{V}_0 = \boldsymbol{V}_{0|0}$，最优滤波估计 $\hat{\boldsymbol{Y}}_{t|t}$ 可由下面的递推关系式给出

$$\hat{\boldsymbol{Y}}_{t|t} = \hat{\boldsymbol{Y}}_{t|t-1} + \boldsymbol{K}_t(Z_t - \boldsymbol{H}_t\hat{\boldsymbol{Y}}_{t|t-1}) \tag{5.5.5}$$

其中

$$\boldsymbol{K}_t = \boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'[\boldsymbol{H}_t\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t' + \sigma_N^2]^{-1} \tag{5.5.6}$$

且有

$$\hat{\boldsymbol{Y}}_{t|t-1} = \boldsymbol{\Phi}_t\hat{\boldsymbol{Y}}_{t-1|t-1}\ , \quad \boldsymbol{V}_{t|t-1} = \boldsymbol{\Phi}_t\boldsymbol{V}_{t-1|t-1}\boldsymbol{\Phi}_t' + \Sigma_a \tag{5.5.7}$$

及

$$\begin{aligned}\boldsymbol{V}_{t|t} &= [I - \boldsymbol{K}_t\boldsymbol{H}_t]\boldsymbol{V}_{t|t-1} \\ &= \boldsymbol{V}_{t|t-1} - \boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'[\boldsymbol{H}_t\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t' + \sigma_N^2]^{-1}\boldsymbol{H}_t\boldsymbol{V}_{t|t-1}\end{aligned} \tag{5.5.8}$$

以上各式中， $t = 1, 2, \cdots$ 。

在 (5.5.5) 中令 $a_{t|t-1} = Z_t - \boldsymbol{H}_t\hat{\boldsymbol{Y}}_{t|t-1} \equiv Z_t - \hat{Z}_{t|t-1}$，该量被称为时刻 t(有限样本) 的新息，因为它是由观测值 Z_t 提供的新的信息，而不是从系统既往观测 (有限) 历史值中所能得到的。因子 $\boldsymbol{K}_t$ 称为“Kalman 增益”阵。(5.5.5) 中的滤波方法是一种通过递推的“预报改进”或“修正”方式，这些方程用来描述最小均方预报，其有效性可以很容易地通过“修正”原理来验证。例如，对于线性预报，可按照修正原理验证如下：

$$\begin{aligned}E[\boldsymbol{Y}_t \mid Z_t, \cdots, Z_1] &= E[\boldsymbol{Y}_t \mid Z_t - \hat{Z}_{t|t-1}, Z_{t-1}, \cdots, Z_1] \\ &= E[\boldsymbol{Y}_t \mid Z_{t-1}, \cdots Z_1] + E[\boldsymbol{Y}_t \mid Z_t - \hat{Z}_{t|t-1}]\end{aligned}$$

其中用到 $a_{t|t-1} = Z_t - \hat{Z}_{t|t-1}$ 与 $Z_{t-1}, \cdots, Z_t$ 独立。由 (5.5.5) 可以看出，基于直到 t 时刻的观测对 $\boldsymbol{Y}_t$ 估计等于直到 $t-1$ 时刻对 $\boldsymbol{Y}_t$ 的预报，再用因子 $\boldsymbol{K}_t$ 与新息 $a_{t|t-1}$ 的乘积加以修正。(5.5.6) 中的方程表明 $\boldsymbol{K}_t$ 可以解释为 $\boldsymbol{Y}_t$ 对新息 $a_{t|t-1}$ 作回归时的回归系数，其中 $\mathrm{var}[a_{t|t-1}] = \boldsymbol{H}_t\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t' + \sigma_N^2$ 及 $\mathrm{cov}[\boldsymbol{Y}_t, a_{t|t-1}] = \boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'$，

由于 $a_{t|t-1}=\boldsymbol{H}_t(Z_t-\hat{Z}_{t|t-1})+N_t$, 以上各式可直接由 (5.5.4) 得到。因此一般的修正关系为

$$\hat{\boldsymbol{Y}}_{t|t}=\hat{Y}_{t|t-1}+\text{cov}[\hat{\boldsymbol{Y}}_{t|t-1},a_{t|t-1}]\{\text{var}[a_{t|t-1}]\}^{-1}a_{t|t-1}$$

这里 $a_{t|t-1}=Z_t-\hat{Z}_{t|t-1}$, 关系式 (5.5.8) 用来作为对误差协方差阵的修正，它可解释为在新息中可利用新信息的协差阵，而预报关系式 (5.5.7) 可由 (5.5.3) 直接得到。

一般而言，未来状态值的预报可由 $\hat{\boldsymbol{Y}}_{t-l|t}=\boldsymbol{\Phi}_{t+l}\hat{\boldsymbol{Y}}_{t+l-1|t}$, 对 $l=1,2,\cdots$ 得到，其中预报误差协方差阵是由 (5.5.7) 式递推得到，即 $\boldsymbol{V}_{t+l|t}=\boldsymbol{\Phi}_{t+l}\boldsymbol{V}_{t+l-1|t}\boldsymbol{\Phi}'_{t+l}+\Sigma_a$ 。最后，未来观测值 Z_{t+l} 的预报由 $\hat{Z}_{t+l|t}=\boldsymbol{H}_{t+l}\hat{\boldsymbol{Y}}_{t+l|t}$ 得出，由于 $Z_{t+l}=\boldsymbol{H}_{t+l}\boldsymbol{Y}_{t+l}+N_{t+l}$, 预报误差为 $v_{t+l|t}=E[(Z_{t+l}-\hat{Z}_{t+l|t})^2]=\boldsymbol{H}_{t+l}\boldsymbol{V}_{t+l|t}\boldsymbol{H}'_{t+l}+\sigma_N^2$ 。

对于 ARIMA 模型来说具有状态空间表示 (5.5.1)—(5.5.2), 且有 $Z_t=z_t=\boldsymbol{H}\boldsymbol{Y}_t$, 及 $\boldsymbol{H}=[1,0,\cdots,0]$, 这时相应的 Kalman 滤波算法构成了基于数据 $z_t,z_{t-1},\cdots,z_1$ 得出精确有限样本预报的另一种方法, 而 ARIMA 过程未来值的估计会受到在 (5.5.5)—(5.5.8) 中用到的初始条件设定的影响。对于平稳零均值过程 z_t 来说，适宜的初始条件为 $\hat{\boldsymbol{Y}}_{0|0}=\boldsymbol{0}$(零向量), 和 $\boldsymbol{V}_{0|0}=\text{cov}[\boldsymbol{Y}_0]\equiv\boldsymbol{V}_*$ 。而 $\boldsymbol{Y}_0$ 的协方差阵，在平稳性条件下该阵很容易由 $\boldsymbol{Y}_t$ 的定义确定。特别, 状态向量遵从平稳的向量 AR(1) 模型: $\boldsymbol{Y}_t=\boldsymbol{\Phi}\boldsymbol{Y}_{t-1}+\boldsymbol{\Psi}a_t$, 其协方差阵 $\boldsymbol{V}_*=\text{cov}[\boldsymbol{Y}_t]$ 满足 $\boldsymbol{V}_*=\boldsymbol{\Phi}\boldsymbol{V}_*\boldsymbol{\Phi}'+\sigma_a^2\boldsymbol{\Psi}\boldsymbol{\Psi}'$, 很容易由此解出 $\boldsymbol{V}_*$ 。对于非平稳的 AIRMA 过程为了进行说明还需要作另一些假设 (例如见 Ansley 和 Kohn[13],Bell 和 Hillmer[29]) 。随着 t 的增大， Kalman 滤波算法提前 l 期预报 $\hat{z}_{t+l|t}$ 的“稳定”值及预报误差方差 $v_{t+l|t}$ 很快地趋近于节 5.1 和 5.2 所给出的表达式，即 $\hat{z}_t(l)$ 和 $V(l)=\sigma_a^2(1+\sum_{j=1}^{l-1}\psi_j^2)$ 。

特别，对于 ARIMA 过程的状态空间形式，我们可以方便地通过 Kalman 滤波方程 (5.5.5)— (5.5.8) 得出精确 (有限样本) 的提前一期预报 $\tilde{z}_{t-1}(l)=E[z_t|z_{t-1},\cdots,z_1]$ 及其误差方差 $v_t=$

$HV_{t|t-1}H'$ 。上述算法对计算 ARIMA 过程 n 个观测值 $z_1,\cdots,z_n$ 的似然函数非常有用，并可用于模型参数的极大似然估计 (例如可见 Jones[123] 及 Gardner,Harvey 和 Phillips[94]), 在 7.5 节中将简要讨论。

5.6 总结

本章的结果可概述如下：令 $\tilde{z}_t$ 为观测时间序列与任一已知确定性时间函数 $f(t)$ 的离差。特别地，对于平稳序列，$f(t)$ 可能等于序列的均值 μ, 或等于 0, 这时 $\tilde{z}_t$ 就是观测序列。然后，我们考虑一般 ARIMA 模型

$$\phi(B)\nabla^d\tilde{z}_t=\theta(B)a_t$$

或者

$$\varphi(B)\tilde{z}_t=\theta(B)a_t$$

最小均方误差预报　如果已知某时间序列原点 t 之前的信息，$\tilde{z}_{t+l}$ 的最小均方误差预报 $\hat{\tilde{z}}_t(l)$ 就是条件期望

$$\hat{\tilde{z}}(l)=[\tilde{z}_{t+l}]=E[\tilde{z}_{t+l}|\tilde{z}_t,\tilde{z}_{t-1},\cdots]$$

提前 1 期预报误差　必然的结论是：提前 1 期预报误差就是由模型中的 a 产生，它是互不相关的。

预报的计算　通常在实际应用中，直接由差分方程计算预报值是最简单的，由下式给出：

$$\begin{aligned}\hat{\tilde{z}}(l)=\varphi_1[\tilde{z}_{t+l-1}]+\cdots+\varphi_{d+p}[\tilde{z}_{t+l-p-d}]+[a_{t+l}]\\-\theta_1[a_{t+l-q}]-\cdots-\theta_q[a_{t+l-q}]\end{aligned}\tag{5.6.1}$$

在 (5.6.1) 计算过程中，要代入已有 z 的真值，z 的预报值，以及已有 a 的实际值和未来 a 的值 (零) 。当然首先要利用第 6 、7 和第 8 章中给出的方法，得出恰当的模型和参数估计值，然后再进行预报。

预报的概率限　预报的概率限可通过以下步骤获得：

1. 首先由下式计算 ψ 权

$$
\begin{aligned}
&\psi_0 = 1 \\
&\psi_1 = \varphi_1 - \theta_1 \\
&\psi_2 = \varphi_1\psi_1 + \varphi_2 - \theta_2 \\
&\ \vdots \quad \vdots \qquad\quad \vdots \quad \vdots \\
&\psi_j = \varphi_1\psi_{j-1} + \cdots + \varphi_{p+d}\psi_{j-p-d} - \theta_j
\end{aligned}
\tag{5.6.2}
$$

这里， $\theta_j = 0, j > q$ 。

2. 将任何给定的概率水平 ϵ, 以及每一预报提前期 l 代入下式

$$
\widetilde{z}_{t+l}(\pm) = \hat{\widetilde{z}}_t(l) \pm u_{\varepsilon/2}\left(1 + \sum_{j=1}^{l-1}\psi_j^2\right)^{1/2}\sigma_a \tag{5.6.3}
$$

在应用中，白噪声的标准差 σ_a 用它的估计值 s_a 代替， $u_{\varepsilon/2}$ 是离差，超出标准正态分布相应范围的比例是 $\varepsilon/2$ 。

修正预报　如果我们得到了一个新的离差 $\widetilde{z}_{t+1}$, 通过计算新的预报误差 $a_{t+1} = \widetilde{z}_{t+1} - \hat{\widetilde{z}}_t(1)$ 以及利用差分方程 (5.6.1) (用 $t+1$ 代替 t), 这时可把预报的时间原点改为 $t+1$ 。然而另一种修正预报的方法就是利用在时间 t 的预报值 $\hat{\widetilde{z}}_t(1), \hat{\widetilde{z}}_t(2), \cdots, \hat{\widetilde{z}}_t(L-1)$, 由

$$
\hat{\widetilde{z}}_t(l) = \hat{\widetilde{z}}_t(l+1) + \psi_l a_{t+1} \tag{5.6.4}
$$

得出原点 $t+1$ 提前 $L-1$ 期的预报值 $\hat{\widetilde{z}}_{t+1}(1), \hat{\widetilde{z}}_{t+1}(2), \cdots, \hat{\widetilde{z}}_{t+1}(L-1)$ 然后利用差分方程 (5.6.1) 得到最后一个预报值 $\hat{\widetilde{z}}_{t+1}(L)$ 。

表示预报的其它方法　以上讨论的就是具体实现预报所使用的方法，然而，下述另一些预报形式是为了从理论角度考察预报的性质，这些预报形式都是由差分方程得出：

1. 求和形式的预报。 对于 $l > q-p-d$, 预报值都在唯一的曲线

$$\hat{\hat{z}}(l) = b_0^{(t)} f_0(l) + b_1^{(t)} f_1(l) + \cdots + b_{p+d-1}^{(t)} f_{p+d-1}(l) \tag{5.6.5}$$

上，该曲线是由"关键值"$\hat{\hat{z}}_t(q), \hat{\hat{z}}_t(q-1), \cdots, \hat{\hat{z}}_t(q-p-d+1)$ 所决定，其中 $\hat{\hat{z}}_t(-j) = \tilde{z}_{t-j}$ $(j = 0, 1, 2, \cdots)$。若 $q > p+d$, 则前 $q-p-d$ 个预报值不在曲线上。一般，平稳自回归算子生成 (5.6.5) 中的衰减指数和衰减正弦项，而非平稳算子 ∇^d 生成直到 $d-1$ 次的多项式。

在 (5.6.5) 中的自适应系数 $b_j^{(t)}$ 从 t 到 $t+1$ 时刻的修正，可归结为依赖于最后的提前 1 期预报误差 a_{t+1}, 对应的一般公式为

$$\boldsymbol{b}^{(t+1)} = \boldsymbol{L}'\boldsymbol{b}^{(t)} + \boldsymbol{g}a_{t+1} \tag{5.6.6}$$

该式在附录 A5.3 中给出。作为特例，(5.4.5) 和 (5.4.13) 分别给出了 IMA(0,1,1) 和 IMA(0,2,2) 过程自适应系数的修正。

2. 预报作为既往观测值的加权和。 把预报作为既往观测值的加权和，从理论的角度来看，是很具有启发性的。因此，若模型写成逆转形式

$$a_t = \pi(B)\tilde{z}_t = (1 - \pi_1 B - \pi_2 B^2 - \cdots)\tilde{z}_t$$

提前 1 期预报便是

$$\hat{\hat{z}}_t(1) = \pi_1 \tilde{z}_t + \pi_2 \tilde{z}_{t-1} + \cdots \tag{5.6.7}$$

多步预报由下式得到

$$\hat{\hat{z}}_t(l) = \pi_1[\tilde{z}_{t+l-1}] + \pi_2[\tilde{z}_{t+l-2}] + \cdots \tag{5.6.8}$$

在 (5.6.8) 中计算条件期望，若实际值为已知，便用以代之 z, 否则用预报值代替。

另外，任意提前期的预报可以写为有效观测值的线性函数:

$$\hat{\tilde{z}}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} \tilde{z}_{t+j-1}$$

其中 $\pi_j^{(l)}$ 是 π_j 的函数。

常数项在预报中的作用　我们还注意到，在 ARIMA(p,d,q) 模型 $\varphi(B)z_t = \theta_0+\theta(B)a_t$(其中 $\varphi(B) = \phi(B)\nabla^d$) 中非零常数项 θ_0 对预报存在影响。因此，在 (5.3.3) 附加了一个确定性的 d 次多项式项 $(\mu_w/d!)\nabla^d$，这里，$w_t = \nabla^d z_t$ 及 $\mu_w = E[w_t] = \theta_0/(1-\phi_1-\phi_2-\cdots-\phi_p)$. 由此得出，预报值不再满足 (5.3.2) 中的关系式 $\varphi(B)\hat{z}_t(l) = 0$，而是满足 $\varphi(B)\hat{z}_t(l) = \theta_0, l > q$，确定性的 d 次多项式给出了该非齐次差分方程的一个特解。因此，在常数项 θ_0 非零的情形，ARIMA 模型也可以表示为 $\phi(B)(\nabla^d z_t - \mu_w) = \theta(B)a_t, \mu_w \neq 0$，而 (5.6.5) 的预报形式可视为表示 $\tilde{z}_{t+l} = z_{t+l} - f(t+l)$ 的预报，这里，$f(t+l) = (\mu_w/d!)(t+l)^d + g(t+l), g(t)$ 是 t 的不超过 $d-1$ 次确定性多项式 (存在 $g(t) = 0$ 的可能) 。譬如，在 ARIMA 模型中 $d = 1$ 的情形，节 5.4.6 的例子考虑的 ARIMA(1,1,1) 就是这类模型。然而，当 $\theta_0 \neq 0$ 时，最终预报误差为 $\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)}\phi^l$，其中应包含另外的确定性线性趋势项 $\mu_w l$，这里 $\mu_w = \theta_0/(1-\phi)$，这类似于对 AIRMA(1,1,0) 模型的例子 (5.4.17) 得到的结果。应该注意到，对于 z_t 是平稳过程的特殊情形，$d = 0$，这时 (5.3.3) 中的附加确定性项退化为过程 z_t 的均值 $\mu = E[z_t]$ 。

A 5.1　预报误差之间的相关

A5.1.1　不同时间原点预报误差的自相关函数

尽管对于最优预报而言，提前 1 期预报误差是互不相关的，但是对于提前多期预报一般并不成立。考虑分别在时间原点 t 和 $t-j$ 所作提前期为 l 的预报，这里 j 是正整数。因此，如果 $j =$

$l, l+1, l+2, \cdots$，则相应的预报误差中不包含共同的分量，但是，对于 $j=1,2,\cdots,l-1$ 的情形，预报误差中将都含有某些 a 。具体地说便是

$$e_t(l)=z_{t+l}-\hat{z}_t(l)=a_{t+l}+\psi_1 a_{t+l-1}+\cdots+\psi_{l-1}a_{t+1}$$
$$e_{t-j}(l)=z_{t+l-j}-\hat{z}_{t-j}(l)=a_{t-j+l}+\psi_1 a_{t-j+l-1}+\cdots+\psi_{l-1}a_{t-j+1}$$

当 $j<l$ 时，提前 l 预报误差的滞后为 j 的自协方差是

$$E[e_t(l)e_{t-j}(l)]=\sigma_a^2\sum_{i=j}^{l-1}\psi_i\psi_{i-j} \qquad \text{(A5.1.1)}$$

这里 $\psi_0=1$ 。相应的自相关是

$$\rho[e_t(l),e_{t-j}(l)]=\begin{cases}\dfrac{\sum\limits_{i=j}^{l-1}\psi_i\psi_{i-j}}{\sum\limits_{i=0}^{l-1}\psi_i^2}, & 0 < j < l\\ 0, & j > l\end{cases} \qquad \text{(A5.1.2)}$$

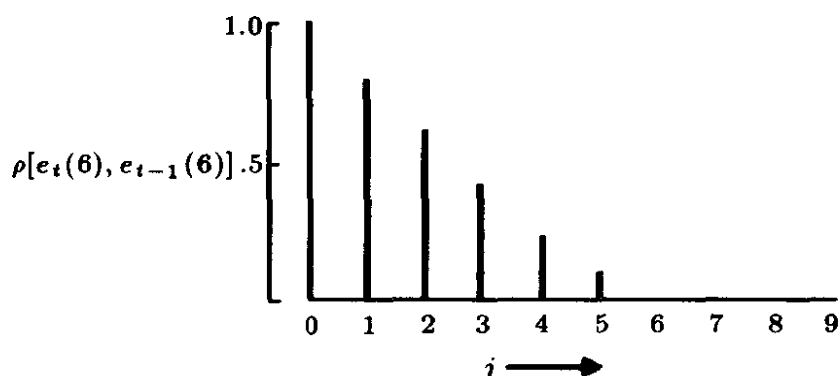
在第 7 章中我们说明了用 (1,1,0) 模型 $(1-0.8B)\nabla z_t=a_t$ 可对图 4.1 中的序列 C 进行很好的拟合。为了解释 (A5.1.2) 式，我们对上述模型计算提前期为 6 的预报误差自相关函数。在节 5.2.2 中给出了该模型的 ψ 权 $\psi_1,\psi_2,\cdots,\psi_5$ 为 1.80,2.44,2.95,3.36,3.69 。因此，例如滞后为 1 的自相关是

$$E[e_t(6)e_{t-1}(6)]=\sigma_a^2[1.80\times 1.00)+(2.44\times 1.80)$$
$$+\cdots+(3.69\times 3.36)]=35.70\sigma_a^2$$

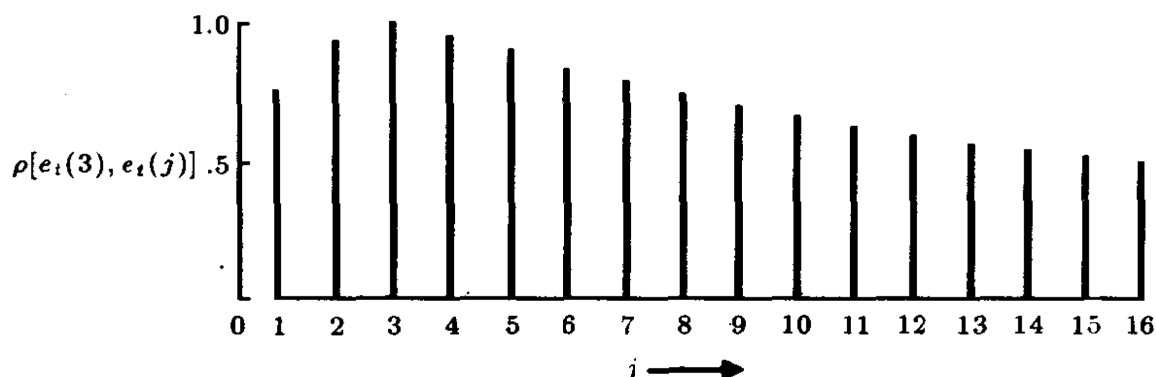
除以 $E[e_t^2(6)]=43.85\sigma_a^2$，我们得到 $\rho[e_t(6),e_{t-1}(6)]=0.81$ 。在表 A5.1 中给出了前 6 个自协方差，并在图 A5.1(a) 中绘出了图形。不出所料，滞后超过 5 的自相关都是零。

表 A5.1　序列 C 步长为 6 的预报误差之间的协方差

j	0	1	2	3	4	5	6
$\rho[e_t(6),e_{t-j}(6)]$	1.00	0.81	0.61	0.41	0.23	0.08	0.00



(a) 序列 C 在不同原点、提前期 $l=6$ 预报误差之间的协方差



(b) 序列 C 在同一原点、提前期为 3 和 j 的预报误差之间的协方差

图 A5.1　序列 C 不同预报误差之间的协方差

A5.1.2　在某一时间原点上不同提前期预报误差之间的相关

假定我们在同一固定的时间原点 t 上得出不同提前期的预报序列，那么，这些预报误差之间是相关的。对于 $j=1,2,3,\cdots$，我们有

$$e_t(l)=z_{t+l}-\hat{z}_t(l)=a_{t+l}+\psi_1 a_{t+l-1}+\cdots+\psi_{l-1}a_{t+1}$$

$$e_t(l+j)=z_{t+l+j}-\hat{z}_t(l+j)=a_{t+l+j}+\psi_1 a_{t+l+j-1}+\cdots+\psi_j a_{t+l}+\psi_{j+1}a_{t+l-1}+\cdots+\psi_{l+j-1}a_{t+1}$$

表 A5.2 在固定原点所做提前 3 期和 j 期预报误差之间的相关 (序列 C)

j	$\rho[e_t(3),e_t(j)]$	j	$\rho[e_t(3),e_t(j)]$
1	0.76	9	0.71
2	0.94	10	0.67
3	1.00	11	0.63
4	0.96	12	0.60
5	0.91	13	0.57
6	0.85	14	0.54
7	0.80	15	0.52
8	0.75	16	0.50

由此得到，在原点 t 提前期为 l 和 $l+j$ 的预报之间的协方差是 $\sigma_a^2\sum_{i=0}^{l-1}\psi_j\psi_{j+i}$, 这里 $\psi_0=1$ 。因此在原点 t 提前期为 l 和 $l+j$ 的预报误差之间的相关系数是

$$\rho[e_t(l),e_t(l+j)]=\frac{\sum_{i=0}^{l-1}\psi_i\psi_{j+i}}{(\sum_{h=0}^{l-1}\psi_h^2\sum_{g=0}^{l+j-1}\psi_g^2)^{1/2}} \tag{A5.1.3}$$

为了说明 (A5.1.3), 我们对序列 C 在同一原点作预报，并计算步长为 3 和步长为 $j=1,2,\cdots,16$ 的预报误差之间的相关。例如，利用 (A5.1.3) 和在节 5.2.2 中给出的 ψ 权，有

$$\begin{aligned}E[e_t(3)e_t(5)] &= \sigma_a^2[(1.00\times 2.44)+(1.80\times 2.95)+(2.44\times 3.36)]\\ &= 15.94\sigma_a^2\end{aligned}$$

表 A5.2 中给出了相对于提前期 $j=1,2,\cdots,16$ 的协方差，并在图 A5.1(b) 中绘出了图形。正如所料，从同一原点所做不同提前期的预报值之间是高度相关的。

A5.2 任意提前期的预报权

在本附录中，对于任意提前期的预报，我们考虑应用于 z 的既往值预报权 $\pi_j^{(l)}$ 的另一种计算方法。为了推导这一结果，我们利用恒等式 (3.1.7), 即

$$(1+\psi_1 B+\psi_2 B^2+\cdots)(1-\pi_1 B-\pi_2 B^2-\cdots)=1$$

由该式我们可以得出 π 权用 ψ 来表示，反之亦然。

令系数相等，我们看到对 $j \geqslant 1$ 有

$$\psi_j=\sum_{i=1}^{j}\psi_{j-i}\pi_i \qquad (\psi_0=1) \tag{A5.2.1}$$

由此有

$$\begin{aligned}
&\psi_1=\pi_1 &&, \quad \pi_1=\psi_1\\
&\psi_2=\psi_1\pi_1+\pi_2 &&, \quad \pi_2=\psi_2-\psi_1\pi_1\\
&\psi_3=\psi_2\pi_1+\psi_1\pi_2+\pi_3 &&, \quad \pi_3=\psi_3-\psi_2\pi_1-\psi_1\pi_2
\end{aligned}$$

由 (5.3.6) 式有

$$\begin{aligned}
\hat{z}_t(l)=\pi_1\hat{z}_t(l-1)+\pi_2\hat{z}_t(l-2)+\cdots+\pi_{l-1}\hat{z}_t(1)\\
+\pi_l z_t+\pi_{l+1}z_{t-1}+\cdots
\end{aligned} \tag{A5.2.2}$$

由于在 (A5.2.2) 中的每一个预报本身都是观测值 $z_t, z_{t-1}, \cdots$ 的函数，因此我们可以写成

$$\hat{z}_t(l)=\pi_1^{(l)}z_t+\pi_2^{(l)}z_{t-1}+\pi_3^{(l)}z_{t-2}+\cdots$$

式中提前期为 l 的预报权可以由提前期为 1 的预报权 $\pi_j^{(1)}=\pi_j$ 算出。现在我们来说明，权 $\pi_j^{(l)}$ 可以利用下面等式得到

$$\pi_j^{(l)}=\sum_{i=1}^{l}\psi_{l-i}\pi_{i+j-1} \tag{A5.2.3}$$

例如，提前 3 期预报的权为

$$\begin{aligned}
\pi_1^{(3)}&=\pi_3+\psi_1\pi_2+\psi_2\pi_1\\
\pi_2^{(3)}&=\pi_4+\psi_1\pi_3+\psi_2\pi_2\\
\pi_2^{(3)}&=\pi_5+\psi_1\pi_4+\psi_2\pi_3
\end{aligned}$$

等等。为了推出 (A5.2.3) 式，我们写成

$$
\begin{aligned}
\hat{z}_t(l) &= \qquad\qquad\qquad\quad \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots \\
\hat{z}_{t+l-1} &= \psi_1 a_{t+l-1} + \cdots + \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots
\end{aligned}
$$

两式相减，我们得到

$$
\hat{z}_t(l) = \hat{z}_{t+l-1}(1) - \psi_1 a_{t+l-1} - \psi_2 a_{t+l-2} - \cdots - \psi_{l-1} a_{t+l}
$$

因此

$$
\begin{aligned}
\hat{z}_t(l) = & \pi_1 z_{t+l-1} + \pi_2 z_{t+l-2} + \cdots + \pi_{l-1} z_{t+1} + \pi_l z_t + \pi_{l+1} z_{t-1} + \cdots \\
& + \psi_1(-z_{t+l-1} + \pi_1 z_{t+l-2} + \cdots + \pi_{l-2} z_{t+1} + \pi_{l-1} z_t + \pi_l z_{t-1} + \cdots) \\
& + \psi_2(-z_{t+l-2} + \cdots + \pi_{l-3} z_{t+1} + \pi_{l-2} z_t + \pi_{l-1} z_{t-1} + \cdots) \\
& + \cdots \\
& + \psi_{l-1}(-z_{t+1} + \pi_1 z_t + \pi_2 z_{t-1} + \cdots)
\end{aligned}
$$

利用关系式 (A5.2.1), 每个 $z_{t+l-1}, \cdots, z_{t+1}$ 的系数都为零，再合并同类项，于是我们便得到所需要的结果 (A5.2.3) 式。另一方面，我们也可以利用如下递推公式

$$
\pi_j^{(l)} = \pi_{j+1}^{(l-1)} + \psi_{l-1} \pi_j \tag{A5.2.4}
$$

为了进一步说明，我们对模型 $\nabla^2 z_t = (1 - 0.9B + 0.5B^2) a_t$ 计算 2 步预报的权系数。方程 (A5.2.4) 给出

$$
\pi_j^{(2)} = \pi_{j+1} + \psi_1 \pi_j
$$

再利用表 5.4 中的权值 $\psi_1 = 1.1$, 我们可得

$$
\begin{aligned}
\pi_1^{(2)} &= \pi_2 + \psi_1 \pi_1 = 0.490 + (1.1)(1.1) = 1.700 \\
\pi_2^{(2)} &= \pi_3 + \psi_1 \pi_2 = -0.109 + (1.1)(0.49) = 0.430
\end{aligned}
$$

等等。前 12 个权数已在表 5.4 中给出。

A 5.3 采用一般求和形式的预报

A5.3.1 得出求和形式的一般方法

在此我们再次强调，在预报的实际计算中差分方程算法是最简单的。下文给出的对求和形式的一般处理，只是为了对已得到预报做进一步讨论。在具体处理时，并不是象节 5.4 的例子中那样求出预报函数的显式，而是将最终预报函数适当地写成含有 $p+d$ 自适应系数的一般形式。随之我们将说明，当 $q>p+d$ 时，最终预报函数应如何修改才能适合前 $q-p-d$ 个预报值。最后我们要说明，当时间原点从 t 到 $t+1$ 时，如何对自适应系数做相应修正。

若当 $j=0,1,2,\cdots$ 时，我们作如下理解： $\hat{z}_t(-j)=z_{t-j}$，则利用节 5.1.2 中对条件期望的讨论，预报值应满足差分方程

$$
\begin{array}{l}
\hat{z}_t(1)-\varphi_1\hat{z}_t(0)-\cdots-\varphi_{p+d}\hat{z}_t(1-p-d)=-\theta_1 a_t-\cdots-\theta_q a_{t-q+1}\\
z_t(2)-\varphi_1\hat{z}_t(1)-\cdots-\varphi_{p+d}\hat{z}_t(2-p-d)=-\theta_2 a_t-\cdots-\theta_q a_{t-q+2}\\
\quad\vdots\qquad\vdots\qquad\qquad\vdots\qquad\qquad\qquad\qquad\vdots\qquad\vdots\\
z_t(q)-\varphi_1\hat{z}_t(q-1)-\cdots-\varphi_{p+d}\hat{z}_t(q-p-d)=-\theta_q a_t\\
z_t(l)-\varphi_1\hat{z}_t(l-1)-\cdots-\varphi_{p+d}\hat{z}_t(l-p-d)=0\qquad (l>q)
\end{array}
\tag{A5.3.1}
$$

最终预报误差是上述最后一个方程的解，可以写成

$$
\hat{z}_t(l)=b_0^{(t)}f_0(l)+b_1^{(t)}f_1(l)+\cdots+b_{p+d-1}^{(t)}f_{p+d-1}(l)=\sum_{i=0}^{p+d-1}b_i^{(t)}f_i(l)\qquad (l>q-p-d)
\tag{A5.3.2}
$$

当 q 小于或等于 $p+d$ 时，最终预报函数将对任意提前期 $l>1$ 给出预报值 $\hat{z}_t(1),\hat{z}_t(2),\hat{z}_t(3),\cdots$ 。

作为 $q<p+d$ 模型的例子，我们假设

$$
(1-B)(1-\sqrt{3}B+B^2)^2 z_t=(1-0.5B)a_t
$$

有 $p+d=5$ 及 $q=1$, 因此

$$(1-B)(1-\sqrt{3}B+B^2)^2\hat{z}_t(l)=0 \qquad (l=2,3,4,\cdots)$$

其中算子 B 作用于 l 而不是 t . 对该差分方程求解便得到预报函数

$$\hat{z}_t(l)=b_0^{(t)}+b_1^{(t)}\cos(\frac{2\pi l}{12})+b_2^{(t)}l\cos(\frac{2\pi l}{12})+b_3^{(t)}\sin(\frac{2\pi l}{12})+b_4^{(t)}\sin(\frac{2\pi l}{12})$$
$$(l=1,2,\cdots)$$

如果 q 大于 $p+d$, 提前期为 $l<q-p-d$ 时, 预报函数将有含 a 的其它项, 即

$$\hat{z}_t(l)=\sum_{i=0}^{p+d-1}b_i^{(t)}f_i(l)+\sum_{i=0}^{j}d_{li}a_{t-i} \qquad (l<q-p-d) \qquad \text{(A5.3.3)}$$

这里 $j=q-p-d-l$, 将 (A5.3.3) 代入 (A5.3.1) 易得 d 。例如考虑如下随机模型

$$\nabla^2 z_t=(1-0.8B+0.5B^2-0.4B^3+0.1B^4)a_t$$

在该模型中, $p+d=2, q=4, q-p-d=2$, 以及 $\varphi_1=2, \varphi_2=-1, \theta_1=0.8, \theta_2=-0.5, \theta_3=0.4, \theta_4=-0.1$ 。利用递推关系式 (5.2.3), 我们得到 $\psi_1=1.2, \psi_2=1.9, \psi_3=2.2, \psi_4=2.6$ 。那么, 由 (A5.3.3) 有

$$\begin{aligned}
\hat{z}_t(1)&=b_0^{(t)}+b_1^{(t)}+d_{10}z_t+d_{11}a_{t-1}\\
\hat{z}_t(2)&=b_0^{(t)}+2b_1^{(t)}+d_{20}a_t\\
\hat{z}_t(l)&=b_0^{(t)}+b_1^{(t)}l \qquad (l>2)
\end{aligned} \qquad \text{(A5.3.4)}$$

利用 (A5.3.1) 给出

$$\hat{z}_t(4)-2\hat{z}_t(3)+\hat{z}_t(2)=0.1a_t$$

再由 (A5.3.4) 有

$$d_{20}a_t = 0.1a_t$$

因此 $d_{20} = 0.1$ 。类似地，由 (A5.3.1),

$$\hat{z}_t(3) - 2\hat{z}_t(2) + \hat{z}_t(1) = -0.4a_t + 0.1a_{t-1}$$

再利用 (A5.3.4),

$$-0.2a_t + d_{10}a_t + d_{11}a_{t-1} = -0.4a_t + 0.1a_{t-1}$$

得

$$d_{10} = -0.2 \qquad d_{11} = 0.1$$

所以预报函数为

$$\begin{aligned}
\hat{z}_t(1) &= b_0^{(t)} + b_1^{(t)} - 0.2a_t + 0.1a_{t-1} \\
\hat{z}_t(2) &= b_0^{(t)} + 2b_1^{(t)} + 0.1a_t \\
\hat{z}_t(l) &= b_0^{(t)} + b_1^{(t)}l \qquad (l > 2)
\end{aligned}$$

A5.3.2 一般求和形式的修正

利用等式 (5.2.5), 并用 t 代替 $t+1$, 便得到系数的修正公式:

$$\hat{z}_t(l) = \hat{z}_{t-1}(l+1) + \psi_t a_t$$

于是，对 $l > q-p-d$ 有

$$\sum_{o=0}^{p+d-1} b_i^{(t)} f_i(l) = \sum_{i=0}^{p+d-1} b_i^{(t-1)} f_i(l+1) + \psi_l a_t \tag{A5.3.5}$$

通过解 $p+d$ 个这样的方程 (对应于 l 的不同的值), 我们就得到所需要的每个系数的修正公式，具体形式如下:

$$b_i^{(t)} = \sum_{j=0}^{p+d-1} L_{ij} b_j^{(t-1)} + g_i a_t$$

注意，预报函数中每一个系数的修正都只依赖于一步预报误差 $a_t = z_t - \hat{z}_{t-1}(1)$ 。

A5.3.3 与折扣最小平方方法的比较

尽管从计算预报的角度来看，采用求和形式是复杂而又不必要的方法，但是这样做使我们能把目前的最小均方误差预报与另一些类型的预报做比较，而后者已受到广泛的关注。我们记

$$\boldsymbol{F}_j = \begin{bmatrix} f_0(l) & f_1(l) & \cdots f_{p+d-1}(l) \\ f_0(l+1) & f_1(l+1) & \cdots f_{p+d-1}(l+1) \\ f_0(l+p+d-1) & f_1(l+p+d-1) & \cdots f_{p+d-1}(l+p+d-1) \end{bmatrix}$$

$$\boldsymbol{b}^{(t)} = \begin{bmatrix} b_0^{(t)} \\ b_1^{(t)} \\ b_{p+d-1}^{(t)} \end{bmatrix}, \qquad \boldsymbol{\Psi}_l = \begin{bmatrix} \psi_l \\ \psi_{l+1} \\ \psi_{l+p+d} \end{bmatrix}$$

于是，对于 $l, l+1, \cdots, l+p+d$ 利用 (A5.3.5) 式，我们得到对 $l > q-p-d$ 有

$$\boldsymbol{F}_l \boldsymbol{b}^{(t)} = \boldsymbol{F}_{l+1} \boldsymbol{b}^{(t-1)} + \boldsymbol{\psi}_l a_t$$

由此得到

$$\boldsymbol{b}^{(t)} = (\boldsymbol{F}_l^{-1} \boldsymbol{F}_{l+1}) \boldsymbol{b}^{(t-1)} + (\boldsymbol{F}_l^{-1} \boldsymbol{\Psi}_l) a_t$$

或

$$\boldsymbol{b}^{(t)} = \boldsymbol{L}' \boldsymbol{b}^{(t-1)} + \boldsymbol{g} a_t \tag{A5.3.6}$$

(A5.3.6) 式与 Brown[64],[65] 的最小平方折扣算法所给出的修正函数具有相同的代数形式。为了进行对比，我们将后一种方法给出的预报误差记为 e_t, 因此 Brown 修正公式可写成

$$\boldsymbol{\beta}^{(t)} = \boldsymbol{L}' \boldsymbol{\beta}^{(t-1)} + \boldsymbol{h} e_t \tag{A5.3.7}$$

这里， $\boldsymbol{\beta}^{(t)}$ 是修正系数向量。在 (A5.3.6) 和 (A5.3.7) 中出现了同一矩阵，这是必然的，因为该因子只有当时间原点转移、相应系

数改变时才出现，并且会出现在同样的任何公式中。例如考虑直线预报函数

$$\hat{z}_{t-1}(l) = b_0^{(t-1)} + b_1^{(t-1)} l$$

这里 $b_0^{(t-1)}$ 是在 $t-1$ 时刻的纵坐标，即是预报的原点。该式可等价地写成

$$\hat{z}_{t-1}(l) = (b_0^{(t-1)} + b_1^{(t-1)}) + b_1^{(t-1)}(l-1)$$

现在 $(b_0^{(t-1)} + b_1^{(t-1)})$ 是在原点 t 的纵坐标。显然，如果我们把预报值修正成在原点 t 的，尽管预报函数的仍不变，但系数 b_0 必定要做适当的调整。

一般，矩阵 $\boldsymbol{L}$ 并不改变预报函数，只不过是将预报函数重新排列。实际上的修正是对系数向量 $\boldsymbol{g}$ 和 $\boldsymbol{h}$ 进行的。我们将会看到，由最小均方误差预报得出的系数 $\boldsymbol{g}$ 和 Brown 给出的系数 $\boldsymbol{h}$ 是完全不同的。

预报的 Brown 方法

1. 预报函数是从多项式、指数、正弦和余弦的线性组和及乘积的通用函数类中加以选择。

2. 对选定的预报函数用“折扣最小二乘”法对既往值进行拟合。在这种方法中，对系数进行估计和修正，以使加权离差平方和

$$S_{\omega} = \sum_{j=0}^{\infty} \omega_j [z_{t-j} - \hat{\hat{z}}_z(-j)]^2 \qquad \text{(A5.3.8)}$$

达到极小，上述离差即是序列的历史值与利用预报函数在相应历史时刻得出的值。权函数按几何衰减，可任意选择。因此有 $\omega_j = (1-\alpha)^j$，这里的常数 α 通常称为平滑常数，它也是任意选取的，可令其等于 0.1 到 0.3 范围的某个值。

最小均方误差预报和 Brown 预报之间的区别　为了说明这一问题，我们考虑 IBM 股票价格的预报， Brown 曾对此进行

讨论 ([64,p.141]) 。在有关研究中，Brown 采用了一个二次模型，用本书的符号表示就是

$$\hat{\hat{z}}_t(l)=\beta_0^{(t)}+\beta_1^{(t)}l+\frac{1}{2}\beta_2^{(t)}l^2$$

他利用该模型，并运用折扣最小二乘法，提前 3 天对股票价格进行预报。图 A5.2 给出了 IBM 序列中的一段，并给出 Brow 方法得到的结果及与最小均方误差预报的对比。

对于折扣最小二乘法可从以下几个方面加以评论：

1. 预报函数的特性应由随机模型中的自回归算子 $\varphi(B)$ 所决定，而不应是任意的。特别是，依靠对时间序列的直觉来确定是不可靠的。例如，考虑在图 A5.2 中绘出的 IBM 股票价格，可以看到，二次函数可以用来拟合该序列较短的数据段，所做的估算是有效的. 如果这种拟合对于预报来说是恰当的，正如 Brown 所归纳的，得到结论是： 2 次多项式的预报函数是适宜的。

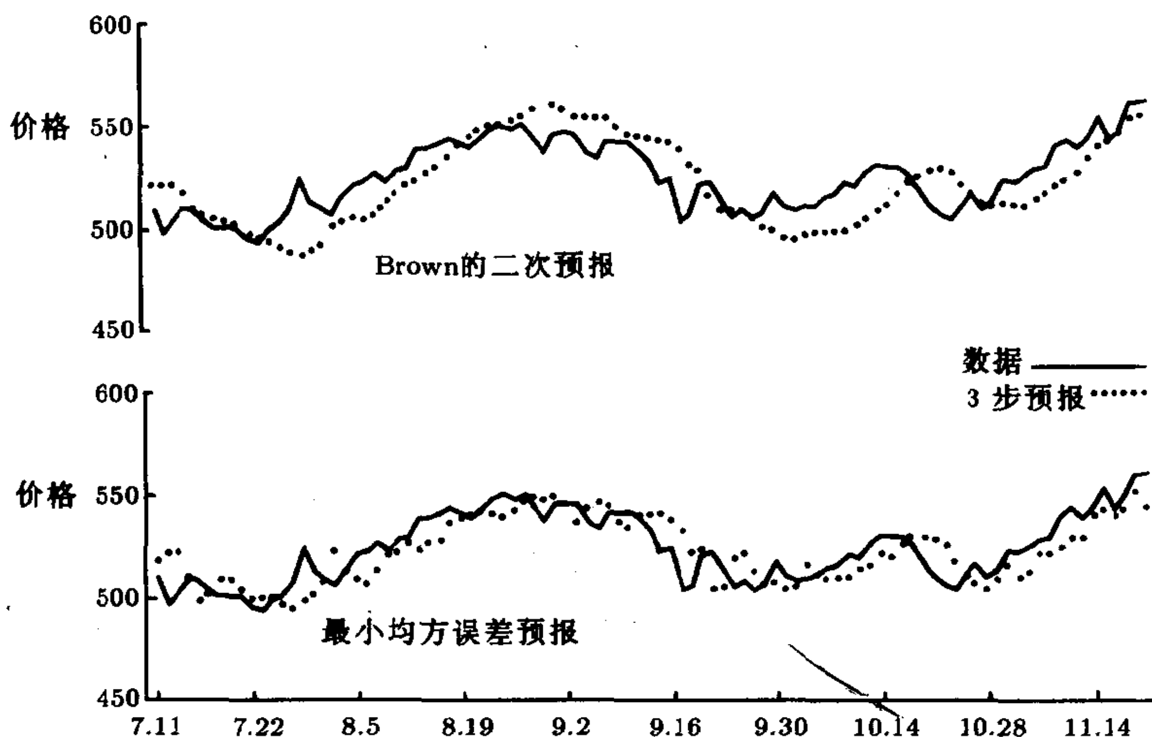


图 A5.2 IBM 股票价格序列与由最佳 IMA(0,1,1) 过程及 Brown 二次预报得到的提前 3 期预报值之间的比较 (数据段从 1960 年 7 月 11 日开始)

对于最一般的线性过程而言，用一个二次函数能得到每一预报步长 $l=1,2,\cdots$ 的最小均方预报，该过程可以定义为 (0,3,3) 模型

$$\nabla^3 z_t = (1-\theta_1-\theta_2 B^2-\theta_3 B^3)a_t$$

正如在节 4.3.3 中所讨论的，上式可以写成

$$\nabla^3 z_t = \nabla^3 a_t + \lambda_0 \nabla^2 a_{t-1} + \lambda_1 \nabla a_{t-1} + \lambda_2 a_{t-1}$$

然而，在第 7 章中我们将要说明，如果用该模型去拟合是正确的，那么参数的最小二乘估计就是 $\lambda_1=\lambda_2=0$，及 $\lambda_0=1.0$ 。因此 $\nabla z_t=(1-\theta B)a_t$ (其中 $\theta=1-\lambda_0$ 接近于零) 是适合的模型，适合的预报多项式是 $\hat{z}_t(l)=\beta_0^{(t)}$，这是 l 的 0 次多项式，而不是 2 次多项式。

2. 在 (A5.3.8) 中权函数 ω_j 的选择必须相应地由随机模型来确定，而不能是任意的。当加上许多限制条件时，使用折扣最小二乘拟合算法将得出最小均方误差预报。这些限制是：

(a) 过程为 (0,1,1) 阶，即 $\nabla z_t(1-\theta B)a_t$ 。

(b) 用 0 次多项式拟合。

(c) 平滑常数 α 令其等于我们所用的 $\lambda=1-\theta$ 。

在本例中，尽管选择了合适的零阶多项式模型，Brown 实际上正是这样做的，且选 $\alpha+\lambda=0.1$ 。但是这种做法是非常不恰当的，因为对于该序列来说，正确的 λ 值应接近于 1.

3. 指数折扣加权最小二乘法要求修正向量 $\boldsymbol{h}$ 中的所有 $p+d$ 个系数都是平滑参数 α 的函数。然而事实上，它们应是 $p+d$ 个独立参数 $(\boldsymbol{\varphi},\boldsymbol{\theta})$ 的函数。

因此，这两种方法之间的差别不是无关紧要的，比较它们对于 IBM 数据所表现的特性是十分有趣的。最小均方误差预报值是 $\hat{z}_t(l)=b_0^{(t)}$，用 $b_0^{(t)}=b_0^{(t-1)}+\lambda a_t$ 来做修正，其中 $\lambda\simeq 1.0$ 。如果取 λ 严格等于 1，这就相当于

$$\hat{z}_t(l)=z_t$$

表 A5.3　两种方法的预报均方误差的对比 — 用最佳 IMA(0,1,1) 过程得到的 3 步预报和 Brown 折扣预报

	滞后时间									
	1	2	3	4	5	6	7	8	9	10
MSE(Brown)	102	158	218	256	363	452	554	669	799	944
MSE(λ=0.9)	42	91	136	180	282	266	317	371	427	483

这意味着对于所有未来时刻股票价格的最佳预报值就是它的当前价格[①]。对于股票特性所作的这种假定当然不是什么新东西，只是再现了 Bachelier[17] 的设想。由于 $z_t = Sa_t$，当 $\lambda = 1$ 时蕴含 z_t 是随机游动。

为了将最小均方误差 (MSE) 预报与 Brown 的折扣预报做对比，我们用 1960 年 7 月 11 日至 1961 年 2 月 10 日的 IBM 股票价格序列 (共 150 个观测值) 进行比较。对于这段序列，最小 MSE 预报是用模型 $\nabla z_t = a_t - \theta a_{t-1}$ 得到的，其中 $\theta = 0.1$(或 $\lambda = 1 - \theta = 0.9$)。图 A5.2 给出了最小 MSE 方法的 3 步预报值，同时给出了 Brown 折扣预报的相应值。可以看出，最小 MSE 预报实际上相当于是用当天的股票价格去预报 3 天后的价格，其效果比用较复杂的 Brown 方法得到的结果要好得多。

在表 A5.3 中给出了两种预报不同提前期的均方误差，是直接比较序列值和相应的提前 l 期预报值而算出的。可以看到，Brown 折扣预报的均方误差远远大于最小均方误差预报。

[①] 这一结果接近于真实的假设：除了序列本身的历史值以外，没有适当的信息可以利用，因此只能考虑很短的预报期。对于较长的预报期，增长和转折因素将变得十分重要。

第二部分
随机模型的建立

我们已经看到，阶数为 (p, d, q) 的 ARIMA 过程提供了这样的一类模型，它们所描述的时间序列虽然是非平稳的，但在统计均衡方面是同质的。

AIRMA 过程定义为如下方程

$$\phi(B)(1-B)^d z_t = \theta_0 + \theta(B)a_t$$

这里 $\phi(B)$ 和 $\theta(B)$ 分别是 B 的 p 次和 q 次算子，它们的零点都在单维圆外. 我们已经指出，该模型是很一般的，它包括自回归模型，滑动平均模型，自回归滑动平均混合模型，以及所有三种模型的求和形式。

模型建立的反复方法

通过识别、估计和诊断检验这三个步骤的反复过程，通常都能把这类模型与数据很好地联系起来。

1. 所谓识别，我们所指的是：利用数据及由序列生成的一切信息给出有价值采用的、参数简约的模型子类。

2. 所谓估计，我们是指：在采用模型是恰当的条件下，有效地利用数据对模型参数进行推断。

3. 所谓诊断检验，所指的是：检验拟合模型与数据的关系，以此来揭示模型的不当之处，并实现对模型的改进。

在下面的第 6 章中，我们讨论模型的检验，第 7 章讨论估计，第 8 章讨论诊断和检验。最后在第 9 章中，通过应用于季节性时间序列建模来对所有这些技巧加以说明。

6 模型识别

在这一章中，我们将讨论非季节时间序列模型的识别方法。识别方法是一种粗略的过程，将此方法运用于一组数据上，以期指明值得做进一步研究的那些模型的类型。在这里的具体目的就是要得到在一般的线性 ARIMA 模型中所需 p, d 和 q 值的一些信息，并得出参数的初始估计。这样便得到试探性的模型，这为第 7 章中所述更正式、更有效估计方法的应用提供了起点。

6.1 识别的目的

首先应该说明，识别和估计过程二者必然会交叉，因此，我们可以先去估计一个比我们希望找到的更为详尽模型的参数，然后再决定哪些方面有可能简化。这里我们用估计过程去完成一部分的识别工作。需要说明的是：识别必然是不精确的，这是由于在实际中发生了哪种类型的模型，是在什么样的情况下发生的，这一问题是物理行为的特性，不可能通过纯数学的讨论来决定。而且，在识别阶段对有关问题没有精确的公式可利用，所以必须使用统计上的“无效性”方法。在这一阶段，图示方法特别有用，并需据此作出判断。但是我们必须记住，初步识别只是为我们提供有关模型类型的试探性考虑，随之，还要有效地拟合与检验。

6.1.1 识别过程的步骤

模型

$$\phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \tag{6.1.1}$$

可用来表示给定的时间序列，这是一般的 ARIMA 模型族。现在，我们的任务是，从中识别出适合的子类。我们的方法如下：

1. 对 z_t 进行足够多次的差分使之达到平稳，以期将所研究过程简化为自回归滑动平均混合过程

$$\phi(B)w_t = \theta_0 + \theta(B)a_t$$

其中

$$w_t = (1-B)^d z_t = \nabla^d z_t$$

2. 识别所得到的 ARMA 过程。

为了实现 1 和 2，我们的主要工具就是样本自相关和样本偏相关函数。它们不仅用来帮助猜测模型的形式，而且可以得出参数的近似估计。这些近似值在估计阶段常常是很有用的，它们为这一阶段所用的迭代过程提供了初始值。

6.2 识别技巧

6.2.1 自相关和偏相关函数在识别中的应用

差分阶数的识别　我们在节 3.4.2 中已经看到，对于一个平稳的 $(p,0,q)$ 阶自回归滑动平均混合过程 $\phi(B)\tilde{z}_t = \theta(B)a_t$，其自相关函数满足差分方程

$$\phi(B)\rho_k = 0 \qquad (k > q)$$

另外，若 $\phi(B) = \prod_{i=1}^{p}(1-G_kB)$，假设根互不相同，那么对于第 k 个自相关函数而言，该差分方程解的形式为

$$\rho_k = A_1G_1^k + A_2G_2^k + \cdots + A_pG_p^k \qquad (k > q-p) \tag{6.2.1}$$

平稳性要求 $\phi(B)$ 的零点在单位圆外，即根 $G_1, G_2, \cdots, G_p$ 在单位圆外。

通过考察 (6.2.1) 表明，在平稳模型情形，如果没有一个根接近单位圆的边界，那么，当 k 充分大时自相关函数将很快地"衰减"。现在假设有一个实的单根，比如 G_1, 接近于 1, 有

$$G_1 = 1 - \delta$$

这里，δ 是某个很小的正数，于是，当 k 很大时有

$$\rho_k \simeq A_1(1 - k\delta)$$

自相关函数不是很快地衰，而是缓慢且非常接近于线性地下降。如果不止一个根接近于 1, 则会得到相似的结论.

从而，若自相关函数不是很快地衰减，这一趋向可以作为可能存在接近于 1 的根的标志。被估计的自相关函数通常趋于服从理论自相关函数的特性，因此，如果估计出的自相关函数不是很快地衰减，则能够从逻辑上认为我们可以将基本随机过程 z_t 处理为非平稳的，但可能 ∇z_t 或更高阶的差分是平稳的,

以上表明，被估计的自相关函数不是很快衰减，则认为序列是非平稳的。这并不要求被估计的相关在低阶滞后时不会有很高的相关值。在附录 A6.1 中对此做了详细说明，并对非平稳的 (0,1,1) 过程 $\nabla z_t = (1 - \theta B)a_t$ 考虑了被估计自相关函数的期望特征。期望值之比 $E[c_k]/E[c_0]$ 只是缓慢地下降，但是一开始将会依赖于 θ 和序列中观测值的数目，且当 θ 接近于 1 时，该比值也不一定趋于 1. 我们将在 6.3.4 节中对图 4.1 中的序列 A 再次详细地说明这一点。

鉴于以上给出的原因，使得 $w_t = \nabla^d z_t$ 的自相关函数很快衰减，相应的 d 就设定为达到平稳所需的的差分阶数。在实际中，d 通常为 0,1 或 2, 并且通常考察原序列、一阶或二阶差分序列自相关函数估计值的前 20 个就够了。

对得到平稳 ARMA 过程的识别　试探性地确定出 d 之后，我们接下来研究经适当差分后的序列 $x_t = (1-B)^d z_t$，其被估计的自相关和偏相关函数的一般性状，以此为确定自回归和滑动平均算子的阶数 p 和 q 提供线索。在操作时，我们要回忆一下在第 3 章中曾讨论过的内容：有关滑动平均、自回归和混合过程它们的理论自相关和偏相关函数的特征行为。

简要地说，p 阶自回归过程的自相关是拖尾的，而它的偏相关函数在滞后 p 步之后是截尾的。与此相应，q 阶滑动平均过程的自相关函数在滞后 q 步之后是截尾的，而它的偏相关函数是拖尾的。进一步，对于一个包含 p 阶自回归和 q 阶滑动平均的混合过程，其自相关函数在 $q-p$ 步滞后之后是混合的指数和衰减正弦波。与此相应，混合过程的偏相关函数在 $p-q$ 步滞后之后被混合的指数和衰减正弦所控制 (见表 3.3)

一般地，用自相关函数度量的自回归 (滑动平均) 特性趋同于用偏相关函数度量的滑动平均 (自回归) 特性。例如，一阶自回归过程的自相关函数呈指数衰减，而它的偏相关函数在一步滞后之后截尾。与此对应，一阶滑动平均过程其自相关函数在一步滞后之后截尾。虽然偏相函数不是精确的指数形式，但被指数所控制，并具有一般的指数特性。

特别重要的是一阶、二阶的自回归和滑动平均过程，以及简单的混合 $(1,d,1)$ 过程。在表 6.1 中总结了这些过程的理论自相关和偏相关函数的性质，该表提供了一个方便的参考表，需要仔细加以研究。读者还应再次参照图 3.2,3.9 和 3.11, 在那些图中给出了二阶自回归过程、二阶滑动平均过程及简单混合 ARMA(1,1) 过程的典型特征。

自相关的估计值与理论值之间的关系　自相关的估计值可能有较大的方差，且彼此之间可能高度相关。正因如此，如 Kendall [126] 所强调的，不可能指望自相关函数的估计值与理论值十分贴近。特别是，当理论自相关函数已经衰减了，而估计的自相关函

表 6.1 经 d 阶差分 (p,d,q) 阶 ARIMA 过程自相关函数的特征 *

	阶		数		
	$(1,d,0)$	$(0,d,1)$	$(2,d,0)$	$(0,d,2)$	$(1,d,1)$
ρ_k 的特征	指数衰减	只有 ρ_1 非零	指数或衰减正弦波的混合	只有 ρ_1 和 ρ_2 非零	自一步滞后之后指数衰减
ϕ_{kk} 的特征	只有 ϕ_{11} 非零	指数控制衰减	只有 ϕ_{11} 和 ϕ_{22} 非零	指数或衰减正弦波的混合	自一步滞后起被指数衰减控制
初始估计	$\phi_1=\rho_1$	$\rho_1=\dfrac{-\theta_1}{1+\theta_1^2}$	$\phi_1=\dfrac{\rho_1(1-\rho_2)}{1-\rho_1^2}$ $\phi_2=\dfrac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_1=\dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$ $\rho_2=\dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$\rho_1=\dfrac{(1-\theta_1\phi_1)(\phi_1-\theta_1}{1+\theta_1^2-2\phi_1\theta_1}$ $\rho_2=\rho_1\phi_1$
允许域	$-1<\phi_1<1$	$-1<\theta_1<1$	$-1<\phi_2<1$ $\phi_2+\phi_1<1$ $\phi_2-\phi_1<1$	$-1<\theta_2<1$ $\theta_2+\theta_1<1$ $\theta_2-\theta_1<1$	$-1<\phi_1<1$ $-1<\theta_1<1$

* 本书后面的表 A 和图 B,C 和 D 可以简化一阶滑动平均，二阶自回归，二阶滑动平均及 ARMA(1,1) 过程参数近似估计值的计算.

数还可能出现相当大，并发生明显的波动和趋势，这些在理论函数中是没有依据的。在运用估计的自相关函数作为识别工具时，通常能对大致特性有相当的把握，更精细的特征则可能未必代表实际结果，因此，可能需要将两个或更多的模型纳入，以便在建模的估计和检验诊断阶段作进一步的研究。

6.2.2 自相关和偏相关估计的标准差

由于我们不知道理论的相关值，而且我们计算的估计值总会和相应的理论值有所偏差，因此指出一个估计值偏离相应的理论值有多远是很重要的。具体地说，我们需要一些方法来判断在某些特定的滞后 p 或 q 之后，自相关或偏相关是否分别有效地为零。对较大的滞后，在 q 阶滑动平均的假设之下，我们用样本估计值代替理论自相关，可以从 Bartlett 公式 (2.1.13) 的简化形式计算出估计自相关的标准差。于是有

$$\hat{\sigma}[r_k] = \frac{1}{n^{1/2}}[1 + 2(r_1^2 + r_2^2 + \cdots + r_q^2)]^{1/2} \quad (k > q) \qquad (6.2.2)$$

对于偏相关，我们利用 (3.2.36) 引用的结果，在过程为 p 阶自回归的假设之下，$p+1$ 或更高阶偏相关估计值的标准差是

$$\hat{\sigma}[\hat{\phi}_{kk}] \simeq \frac{1}{n^{1/2}} \qquad (k > p) \qquad (6.2.3)$$

正如 Anderson[6] 所指出，对于适当大小的 n, 自相关系数的估计值 (其理论值为零) 的分布近似为正态。因此，在理论自相关系数 ρ_k 为零的假设下，估计值 r_k 除以它的标准差将近似服从单位正态偏差的分布。对于偏相关系数有类似的结果。这些事实可以提供一种非正式的标准，用来指示当滞后超出某特定值后理论自相关和偏相关是否实质上为零。对于正态分布来说，记住下面的事实就足够了：在任意方向上超出一个标准差的偏离概率大约是 1/3, 在任意方向上超出两个标准差的偏离概率大约是 1/20 。

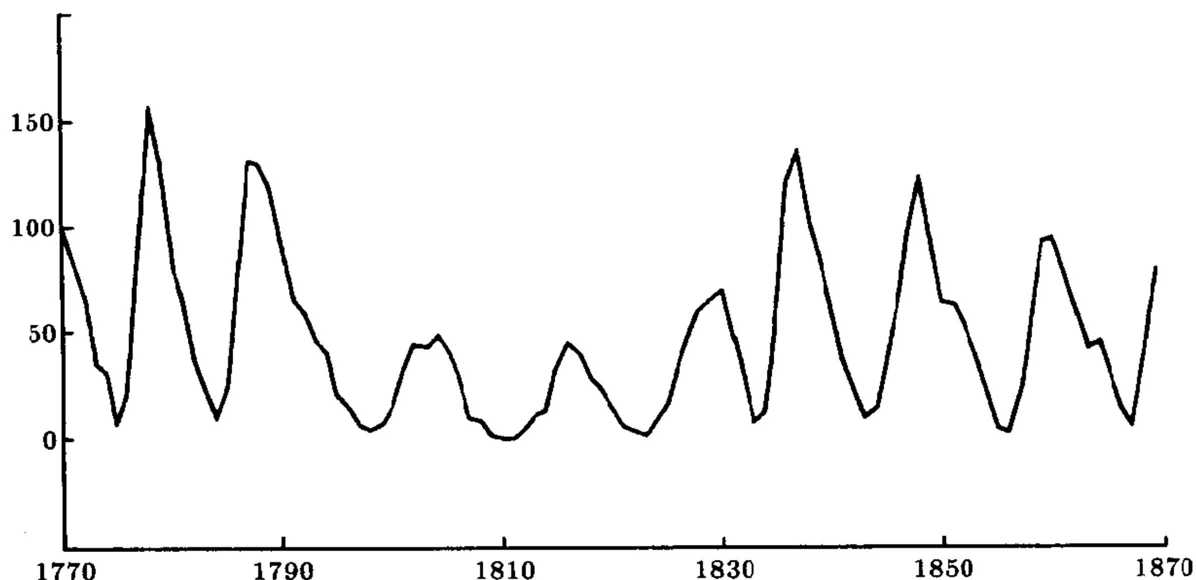


图 6.1 序列 E: 年度 Wö lfer 太阳黑子数 (1770–1869)

6.2.3 一些实际时间序列的识别

在本节中，前面涉及的技巧被应用于 6 个时间序列，记作 A,B,···,F 。序列 A 到 D 在图 4.1 中给出，序列 E 如图 6.1, 序列 F 如图 2.1. 所有这些序列的数据都列在第五部分的“时间序列汇集”中。序列 A,B,C,D 已经在第 4 章作了描述，序列 F 见第 2 章。序列 E 是年度 Wölfer 太阳黑子数的序列，用来度量每年太阳黑子的平均数．正如我们在第 4 章中所指出的，可以料想，序列 A,C 和 D 具有非平稳的特点，这是因为它们体现了某些过程的输出为“不可控”的特征。与此类似，我们可以想象，股票序列 B 将是没有固定的水平。另一方面，我们预料序列 F 将是平稳的，因为它表现的是在同一控制条件下，生产过程产出的变化。类似地，我们预料太阳黑子序列将在长时期保持均衡状态。

有关序列 A 到 F 的 $z,\nabla z$ 和 $\nabla^2 z$ 的自相关估计值在表 6.2 中列出。表 6.3 给出了相应的偏相关估计值。画出相关函数图形对于理解和说明问题很有帮助，在图 6.2 和 6.3 中绘出了序列 A 的自相关和偏向关图形，在图 6.4 和 6.5 中是序列 C 的有关图形。

表 6.2 序列 A 到 F 的自相关估计值

	滞后	自相关 1	2	3	4	5	6	7	8	9	10
		序列 A: 化工过程浓度的每隔 2 小时读数, 197 个观测值									
z	1–10	0.57	0.50	0.40	0.36	0.33	0.35	0.39	0.32	0.30	0.26
	11–20	0.19	0.16	0.20	0.24	0.14	0.18	0.20	0.20	0.14	0.18
∇z	1–10	–0.41	0.02	–0.07	–0.01	–0.07	–0.02	0.15	–0.07	0.04	0.02
	11–20	–0.05	–0.06	–0.01	0.16	–0.17	0.03	0.01	0.08	–0.12	0.15
$\nabla^2 z$	1–10	–0.65	0.18	–0.04	0.04	–0.04	–0.04	0.13	–0.11	0.04	0.02
	11–20	–0.02	–0.02	–0.04	0.18	–0.19	0.08	–0.03	0.09	–0.17	0.20
		序列 B: IMB 普通股日收盘价, 1961.5.17—1962.11.2, 369 个观测值									
z	1–10	0.99	0.99	0.98	0.97	0.96	0.96	0.95	0.94	0.93	0.92
	11–20	0.91	0.91	0.90	0.89	0.88	0.87	0.86	0.85	0.84	0.83
∇z	1–10	0.09	0.00	–0.05	–0.04	–0.02	0.12	0.07	0.04	–0.07	0.02
	11–20	0.08	0.05	–0.05	0.07	–0.07	0.12	0.12	0.05	0.05	0.07
$\nabla^2 z$	1–10	–0.45	–0.02	–0.04	0.00	–0.07	0.11	–0.01	0.04	–0.10	0.02
	11–20	0.04	0.04	–0.12	0.13	–0.17	0.10	0.05	–0.04	–0.01	0.09
		序列 C: 化工生产每分钟温度读数, 226 个观测值									
z	1–10	0.98	0.94	0.90	0.85	0.80	0.75	0.69	0.64	0.58	0.52
	11–20	0.47	0.41	0.36	0.30	0.25	0.20	0.15	0.10	0.05	0.00
∇z	1–10	0.80	0.65	0.53	0.44	0.38	0.32	0.26	0.19	0.14	0.14
	11–20	0.10	0.09	0.07	0.07	0.07	0.07	0.09	0.05	0.04	0.04
$\nabla^2 z$	1–10	–0.08	–0.07	–0.12	–0.06	0.01	–0.02	0.05	–0.05	–0.12	0.12
	11–20	–0.12	0.07	–0.08	0.03	–0.01	–0.06	0.17	–0.10	–0.01	–0.02

表 6.2 序列 A 到 F 的自相关估计值 (续)

	滞后	自相关									
		1	2	3	4	5	6	7	8	9	10
		序列 D: 化工过程粘度的每分钟读数, 310 个观测值									
z	1–10	0.86	0.74	0.62	0.53	0.46	0.41	0.35	0.31	0.27	0.24
	11–20	0.22	0.20	0.18	0.15	0.14	0.13	0.16	0.19	0.21	0.23
∇z	1–10	–0.05	–0.06	–0.07	–0.08	–0.06	0.00	–0.02	–0.02	–0.03	–0.06
	11–20	–0.01	0.04	0.02	–0.07	–0.03	–0.09	–0.02	0.05	–0.01	0.06
$\nabla^2 z$	1–10	–0.50	0.00	0.00	–0.01	–0.02	0.04	–0.01	0.00	0.01	–0.04
	11–20	0.00	0.04	0.03	–0.06	0.04	–0.06	0.00	0.06	–0.06	0.06
		序列 E: Wölfer 太阳黑子数年度值 100 个观测值									
z	1–10	0.81	0.43	0.07	–0.17	–0.27	–0.21	–0.04	0.16	0.33	0.41
	11–20	0.39	0.29	0.14	0.02	–0.06	–0.10	–0.14	–0.18	–0.17	–0.10
∇z	1–10	0.55	–0.02	–0.30	–0.40	–0.40	–0.33	–0.20	0.04	0.26	0.31
	11–20	0.29	0.16	–0.03	–0.12	–0.10	–0.09	–0.09	–0.12	–0.14	–0.05
$\nabla^2 z$	1–10	0.15	–0.31	–0.20	–0.11	–0.09	–0.02	–0.11	–0.04	0.19	0.05
	11–20	0.13	0.09	–0.10	–0.11	0.04	0.01	0.00	–0.03	–0.10	–0.04
		序列 F: 化工生产过程产量, 70 个观测值									
z	1–10	–0.39	0.30	–0.17	0.07	–0.10	–0.05	0.04	–0.04	–0.01	0.01
	11–20	0.11	–0.07	0.15	0.04	–0.01	0.17	–0.11	0.02	–0.05	0.02
∇z	1–10	–0.74	0.43	–0.27	0.16	–0.10	0.01	0.05	-0.05	0.04	–0.05
	11–20	0.11	–0.16	0.12	–0.01	–0.08	0.16	–0.14	0.08	–0.07	0.03
$\nabla^2 z$	1–10	–0.83	0.54	–0.33	0.21	–0.12	0.03	0.04	–0.06	0.06	–0.07
	11–20	0.12	–0.16	0.11	0.00	–0.10	0.16	–0.15	0.10	–0.07	0.03

表 6.3 序列 A 到 F 的偏相关估计值 *

	滞后	偏相关 1	2	3	4	5	6	7	8	9	10
		序列 A：化工过程浓度的每隔 2 小时读数， 197 个观测值									
z	1–10	0.57	0.25	0.08	0.09	0.07	0.15	0.19	–0.03	0.01	–0.01
	11–20	–0.09	–0.04	0.04	0.08	–0.15	0.06	0.13	0.09	–0.06	0.07
∇z	1–10	–0.41	–0.19	–0.17	–0.14	–0.20	–0.23	0.01	–0.04	–0.01	0.06
	11–20	0.02	–0.07	–0.10	0.13	–0.09	–0.15	–0.11	0.04	–0.08	0.12
$\nabla^2 z$	1–10	–0.66	–0.43	–0.33	–0.23	–0.20	–0.36	–0.23	–0.21	–0.23	–0.16
	11–20	–0.07	–0.04	–0.25	0.00	0.04	–0.02	–0.16	–0.03	–0.22	–0.03
		序列 B：IMB 普通股日收盘价， 1961.5.17—1962.11.2， 369 个观测值									
z	1–10	0.996	–0.09	0.01	0.05	0.02	0.02	–0.12	–0.05	–0.02	0.06
	11–20	–0.05	–0.09	–0.03	0.07	–0.08	0.06	–0.14	–0.10	–0.01	–0.08
∇z	1–10	0.09	–0.01	–0.05	–0.03	–0.02	0.13	0.05	0.02	–0.06	0.05
	11–20	0.09	0.03	–0.08	0.08	–0.06	0.14	0.10	0.00	0.07	0.08
$\nabla^2 z$	1–10	–0.45	–0.28	–0.24	–0.20	–0.29	–0.17	–0.13	–0.03	–0.14	–0.16
	11–20	–0.09	0.02	–0.13	0.01	–0.19	–0.13	–0.03	–0.10	–0.10	0.06
		序列 C：化工生产每分钟温度读数， 226 个观测值									
z	1–10	0.99	–0.81	–0.03	–0.02	–0.10	–0.07	–0.01	–0.03	0.04	–0.04
	11–20	–0.15	0.10	–0.14	0.01	–0.10	–0.02	–0.07	–0.11	0.11	–0.13
∇z	1–10	0.81	–0.01	–0.01	0.06	0.03	–0.03	–0.01	–0.08	0.00	0.10
	11–20	–0.14	0.10	–0.05	0.05	0.02	0.06	0.06	–0.17	0.09	0.00
$\nabla^2 z$	1–10	–0.08	–0.08	–0.14	–0.10	–0.03	–0.05	0.02	–0.06	–0.16	0.09
	11–20	–0.14	0.01	–0.09	–0.02	–0.05	–0.09	0.13	–0.13	–0.03	–0.05

* 是通过用最小二乘法拟合阶数增加的自回归过程而得到的.

表 6.3 序列 A 到 F 的偏相关估计值 (续)

	滞后	偏相关									
		1	2	3	4	5	6	7	8	9	10
		序列 D: 化工过程粘度的每分钟读数， 310 个观测值									
z	1–10	0.86	–0.02	0.00	0.01	0.03	0.03	–0.02	0.01	0.00	0.21
	11–20	0.05	0.01	–0.04	–0.03	0.07	0.04	0.10	0.06	0.00	0.06
∇z	1–10	–0.05	–0.06	–0.07	–0.09	–0.08	–0.03	–0.05	–0.05	–0.05	–0.09
	11–20	–0.05	0.01	–0.01	–0.10	–0.07	–0.13	–0.09	–0.02	–0.08	0.00
$\nabla^2 z$	1–10	–0.50	–0.32	–0.24	–0.20	–0.22	–0.16	–0.14	–0.11	–0.07	–0.12
	11–20	-0.15	–0.12	–0.02	–0.06	–0.01	–0.07	–0.12	–0.06	–0.13	–0.07
		序列 E: Wölfer 太阳黑子数年度值 100 个观测值									
z	1–10	0.81	–0.71	0.21	–0.15	0.10	0.10	0.18	0.23	0.01	0.00
	11–20	0.14	–0.16	0.12	0.03	–0.08	–0.14	–0.06	–0.12	0.00	0.05
∇z	1–10	0.57	–0.48	–0.06	–0.27	–0.22	–0.26	–0.29	–0.05	–0.02	–0.16
	11–20	0.13	–0.15	–0.04	0.06	0.12	0.02	0.07	–0.06	–0.09	–0.06
$\nabla^2 z$	1–10	0.15	–0.35	–0.10	–0.21	–0.16	–0.17	–0.36	–0.26	–0.09	–0.33
	11–20	–0.02	–0.13	–0.20	–0.21	–0.10	–0.13	0.00	0.03	–0.01	–0.08
		序列 F: 批量化工过程产量， 70 个观测值									
z	1–10	–0.40	0.19	0.01	–0.07	–0.07	–0.15	0.05	0.00	–0.10	0.05
	11–20	0.18	–0.05	0.09	0.18	0.01	0.43	0.01	–0.14	0.11	0.18
∇z	1–10	–0.76	–0.32	–0.19	–0.16	–0.09	–0.24	–0.15	–0.06	–0.18	–0.28
	11–20	–0.02	–0.16	–0.24	–0.06	–0.44	–0.02	0.12	–0.12	–0.17	–0.24
$\nabla^2 z$	1–10	–0.83	–0.52	–0.38	–0.33	–0.15	–0.24	–0.26	–0.14	–0.09	–0.31
	11–20	–0.12	–0.09	–0.26	0.08	–0.38	–0.39	–0.07	–0.05	–0.03	–0.30

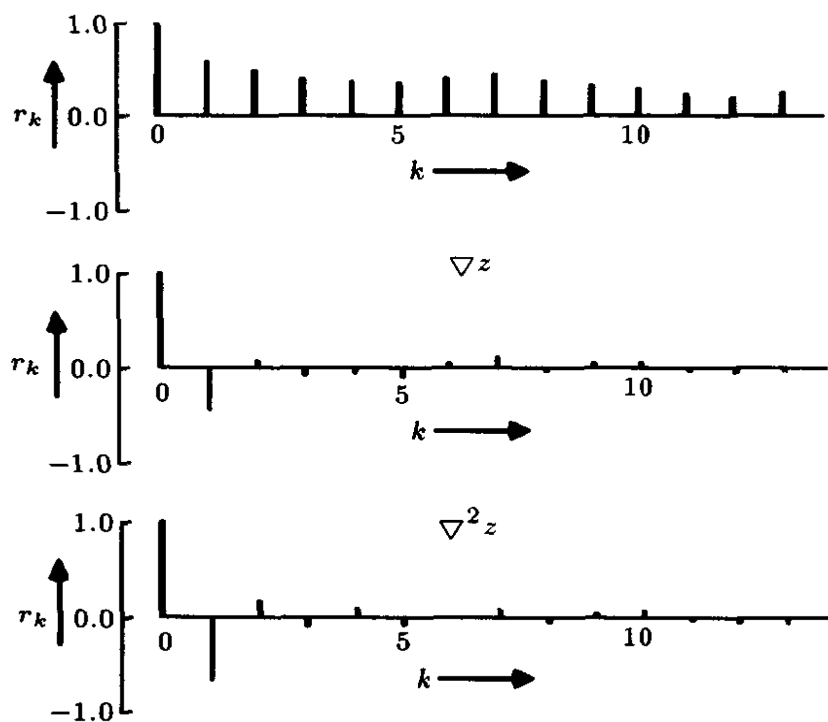


图 6.2　对序列 A 作不同差分的自相关估计

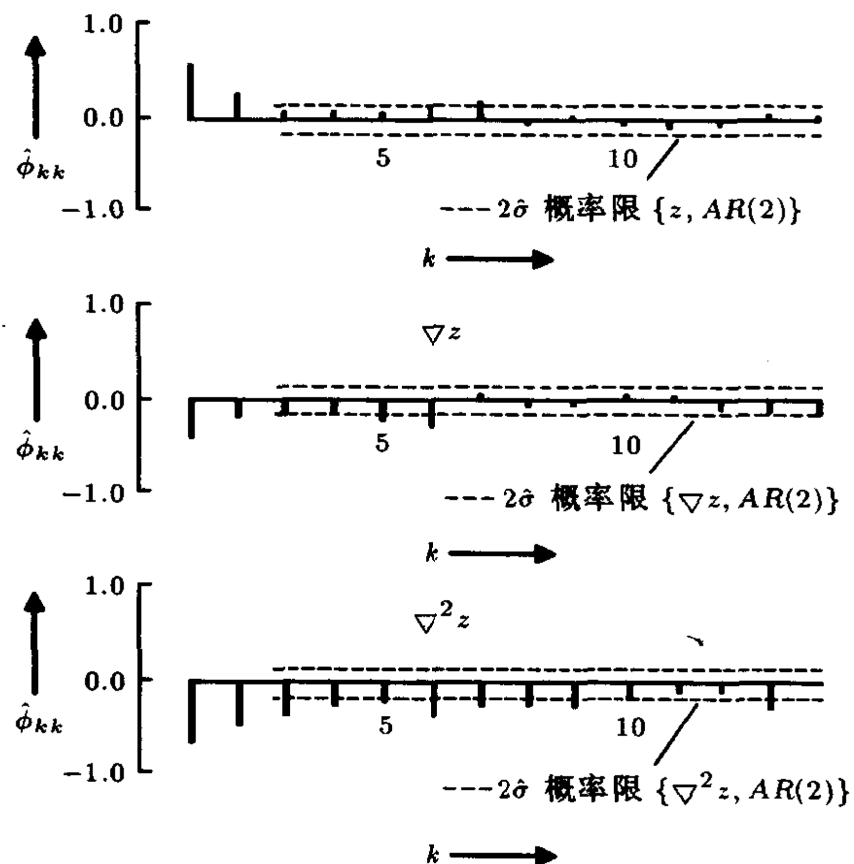


图 6.3　对序列 A 作不同差分的偏相关估计

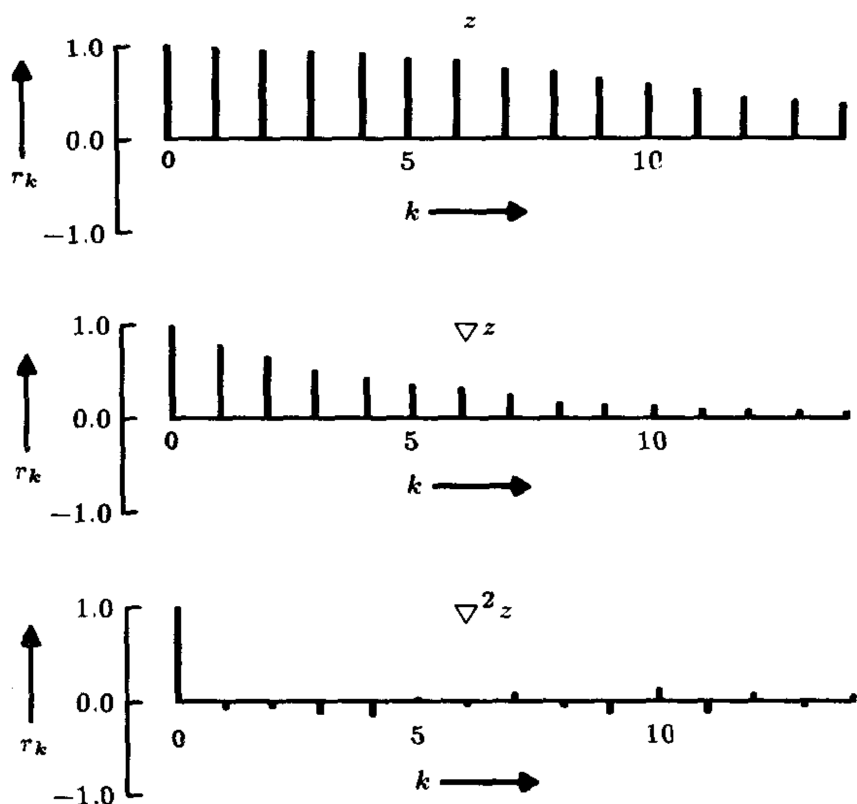


图 6.4 对序列 C 作不同差分的自相关估计

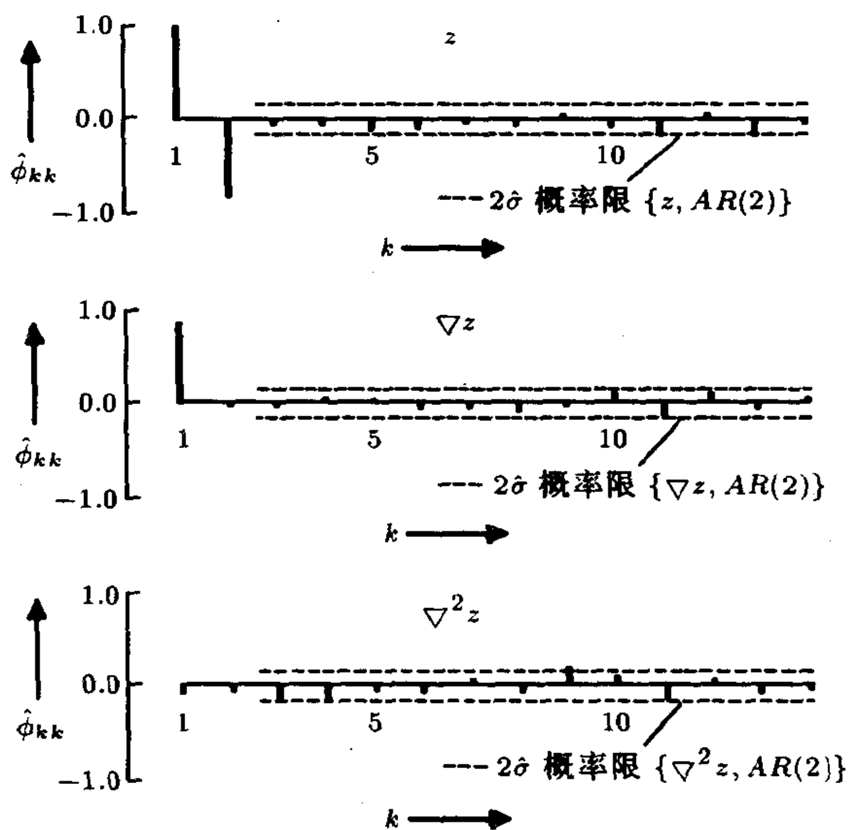


图 6.5 对序列 C 作不同差分的偏相关估计

对于序列 A,∇z 的自相关在一步滞后之后就很小，这表明该时间序列可以用 IMA(0,1,1) 模型来描述。然而，从 z 的自相关函数可以看出，在 1 步滞后之后相关函数非常有规律地衰减。因此，另一种可能是该序列为 (1,0,1) 阶的混合 ARM 过程，z 的偏相关函数也支持这种可能性。以后我们将会看到，这两种可能性实际上得到同一模型。

图 6.4 所示序列 C 的自相关表明，至少一阶差分是必要的。一阶差分后相关函数大致呈指数下降，这使我们联想到自回归参数 ϕ 约为 0.8 的 (1,1,0) 阶过程。另外，我们注意到，$\nabla^2 z$ 的自相关很小，这隐含着序列适合于 IMA(0,2,0) 过程。考虑偏相关函数也能得到相同的结论。对于 z 的考察给出了一般的二阶自回归算子，即关于 ∇z_t 给出了参数 ϕ 约等于 0.8 的一阶自回归过程，有关 $\nabla^2 z_t$ 则给出的是不规则噪声。因此可能的模型是

$$(1-0.8B)(1-B)z_t = a_t$$
$$(1-B)^2 z_t = a_t$$

第二个模型与第一个非常相似，差别仅在于其中一个自回归系数不是 0.8, 而是 1. 在评价相关函数估计值时，画出在 0 上下 $\pm\hat{\sigma}$ 或 $\pm 2\hat{\sigma}$ 的“控制”线是很有帮助的，随之而来存在着心目中设定的标记。图 6.5 边界线上的注释“$\{z, AR(2)\}$ 的 $2\hat{\sigma}$ 界限”是指所使用的 $\hat{\sigma}$ 在过程为二阶自回归这一假设之下是近似正确的。对于 A 到 F 中每个序列的试探性识别在表 6.4 中给出。有关该识别过程还有另外三点需要提到:

1. 对于包含季节分量的序列，使用前面那种简单差分不能得出平稳性。在第 9 章中我们讨论对这类季节性时间序列的适当修正。

2. 如第 4 章中的讨论，在 (6.1.1) 中非零的 θ_0 意味着系统存在 d 阶多项式趋势。对于表 6.4 中的非平稳模型，θ_0 的价值在于能够很好地解释序列的特征。然而，有时在某些实际物理现象中需要考虑这样的分量。但在其它场合则无法确定是否应做这种考

表 6.4　对序列 A 到 F 的模型作试探性检验

序列	差分阶数	差分序列的明显性质	对 z_t 的识别
A	0	混合的 1 阶 AR 和 1 阶 MA	(1,0,1)
	或 1	1 阶 MA	(0,1,1)
B	1	1 阶 MA	(0,1,1)
C	1	1 阶 AR	(1,1,0)
	或 2	不相关噪声	(0,2,2)*
D	0	1 阶 AR	(1,0,0)
	或 1	不相关噪声	(0,1,1)*
E	0	2 阶 AR	(2,0,0)
	或 0	3 阶 AR	(3,0,0)
F	0	2 阶 AR	(2,0,0)

* 滑动平均算子的阶数显然为零，更一般的形式留待后文讨论.

虑。对于在模型中到底是否包含 θ_0, 我们可以在识别阶段从数据得出一些佐证的指标，如通过将 $w_t = \nabla^d z_t$ 的均值 $\overline{w}$ 与它的近似标准差 $\sigma^2(\overline{w}) = n^{-1}\sigma_w^2[1 + 2r_1(w) + 2r_2(w) + \cdots]$ 作比较。

3. 在 3.4.2 中我们注意到，对于 $p - q > 0$ 的任何 ARMA(p, q) 过程, 其自相关函数的整个正半侧是衰减正弦波和指数的混合。当然, 这并不妨碍我们试探性地识别 q, 因为: (a) 偏相关函数在显示出 MA(q) 过程的特性之前, 会出现 p–q 个“异常值”; (b) q 必须可使自相关函数按照一般的形式取初值，即从 ρ_q 回溯到 $\rho_{-(p-q-1)}$ 。

6.2.4　其它模型识别的工具

虽然样本自相关和偏相关函数在模型识别中是非常有用的，但是对于涉及混合模型的情形却不能给出明确的结果。这不能认为是严重的问题，因为正如已强调的那样，模型识别往往是试探性的，并需进一步地考察、诊断检验及必要的修改。然而，在模型识别阶段发展其它的应用工具是很有意义的。这些方法包括由 Gray,Kelley 和 McIntire[97] 提出的 R 和 S 数组方法，Woodward 和 Gray[208] 研究的广义偏相关函数， Cleveland[71] 和 Chatfield[70]

考虑的逆自相关函数，Tsay 和 Tiao[192] 的广义自相关函数，以及由 Akaike[4],Cooper 和 Wood[72], 及 Tsay,Tiao[193] 考察的典型相关分析。

典型相关方法　为了说明起见，我们讨论典型相关分析应用于模型识别的各方面问题。通常，对于维数为 k 和 l(假设 $k < l$) 的两组随机变量 $\boldsymbol{Y}_1 = (y_{11}, y_{12}, \cdots, y_{1k})'$ 和 $\boldsymbol{Y}_2 = (y_{21}, y_{22}, \cdots, y_{2l})'$，典型相关分析涉及确定出线性组合 $U_i = \boldsymbol{a}_i'\boldsymbol{Y}_1$ 和 $V_i = \boldsymbol{b}_i'\boldsymbol{Y}_2, i = 1, 2, \cdots, k$，以及对应的相关系数 $\rho(i) = \text{corr}[U_i, V_i]$，这里 $\rho(1) \geq \rho(2) \geq \cdots \geq \rho(k) \geq 0$，使得 U_i 和 V_j 对于 $i \neq j$ 是互不相关的，且在 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 的所有线性组合中 U_1 和 V_1 具有最大可能的相关 $\rho(1)$，在所有与 U_1,V_1 不相关的 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 的线性组合中，U_2 和 V_2，具有最大可能的相关 $\rho(2)$，依此类推。由此得出的相关系数 $\rho(i)$ 称为 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 之间的典型相关，而变量 U_i 和 V_i 称为相应的典型变量. 如果 $\boldsymbol{\Omega} = \text{cov}[\boldsymbol{Y}]$ 记 $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')$ 的协方差阵，这里 $\boldsymbol{\Omega}_{ij} = \text{cov}[\boldsymbol{Y}_i, \boldsymbol{Y}_j]$，则数值 $\rho^2(i)$ 是矩阵 $\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}$ 的顺序特征值，而满足 $U_i = \boldsymbol{a}_i'\boldsymbol{Y}_1$ 的向量 $\boldsymbol{a}_i$ 是相应的 (标准化) 特征向量，即 $\rho^2(i)$ 和 $\boldsymbol{a}_i'$ 满足

$$[\rho^2(i)\boldsymbol{I} - \boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}]\boldsymbol{a}_t = 0 \quad (i = 1, \cdots, k) \tag{6.2.4}$$

其中 $\rho^2(1) \geq \rho^2(2) \geq \cdots \geq \rho^2(k) \geq 0$(例如，可见 Anderson [7,p.490])。类似地，我们可以定义当给定任何一组变量 $\boldsymbol{Y}_3$ 时，$\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 之间典型偏相关的概念，即是 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 通过对 $\boldsymbol{Y}_3$ 的线性回归，将 $\boldsymbol{Y}_3$ 的影响“调整”掉了，在此之后 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 之间的典型相关，这类似于在节 3.2.5 中讨论的偏相关的定义。应该注意到一个有用的事实：如果存在 (至少)$s < k$ 个与 $\boldsymbol{Y}_2$ 完全不相关、$\boldsymbol{Y}_1$ 的线性独立的线性组合，记作 $\boldsymbol{U} = \boldsymbol{A}'\boldsymbol{Y}_1$，有 $\text{cov}[\boldsymbol{Y}_2, \boldsymbol{U}]$，那么，$\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 之间至少有 s 个零值典型相关系数。这很容易由 (6.2.4) 得出，因为 (至少) 有 s 个独立的特征向量满足 (6.2.4)，且相应有 $\rho(i) = 0$。事实上，零值典型相关函数的数目 $s = k - r$，这里 $r = \text{rank}(\boldsymbol{\Omega}_{21})$。

有关 ARMA 时间序列模型，按照 Tsay 和 Tiao[193] 的方法，我们考虑 $\boldsymbol{Y}_{m,t}=(\widetilde{z}_t,\widetilde{z}_{t-1},\cdots,\widetilde{z}_{t-m})'$, 并检验 $\boldsymbol{Y}_{m,t}$ 和 $\boldsymbol{Y}_{m,t-j-1}=(\widetilde{z}_{t-j-1},\widetilde{z}_{t-j-2},\cdots,\widetilde{z}_{t-j-1-m})'$ 之间的典型相关结构，后者是对 $m=0,1,\cdots$ 和 $j=0,1,\cdots$ 的多种组合。需要回忆的一个重要特征是 ARMA(p,q) 过程 $\widetilde{z}_t$ 的自协方差函数 γ_k 满足 (3.4.2), 且特别有

$$\gamma_k-\sum_{i=1}^{p}\phi_i\gamma_{k-i}=0 \qquad (k>q)$$

因此，例如当 $m\geq p$ 时，(至少) 有一个 $\boldsymbol{Y}_{m,t}$ 的线性组合，使得

$$\widetilde{z}_t-\sum_{i=1}^{p}\phi_i\widetilde{z}_{t-i}=(1,-\phi_1,\cdots,-\phi_p,0,\cdots,0)\boldsymbol{Y}_{m,t}\equiv\boldsymbol{a}'\boldsymbol{Y}_{m,t} \quad (6.2.5)$$

使得

$$\boldsymbol{a}'\boldsymbol{Y}_{m,t}=a_t-\sum_{i=1}^{q}\theta_i a_{t-i}$$

与 $\boldsymbol{Y}_{m,t-j-1}$ 不相关，对所有 $j\geq q$. 于是，特别对于 $m=p,j=q$ 的情形，$\boldsymbol{Y}_{p,t}$ 和 $\boldsymbol{Y}_{p,t-q-1}$ 之间存在一个典型相关，同样在 $\boldsymbol{Y}_{p,t}$ 和 $\boldsymbol{Y}_{p,t-j-1},j>q$ 之间，以及当 $m>p$ 时，$\boldsymbol{Y}_{m,t}$ 和 $\boldsymbol{Y}_{m,t-q-1}$ 之间也是如此，而一般不难证明，对于 $m>p$ 和 $j>q$, 在 $\boldsymbol{Y}_{m,t}$ 和 $\boldsymbol{Y}m,t-j-1,j>q$ 之间存在 $s=\min(m+1-p,j+1-q)$ 个零值典型相关. 由此可以看到，对于 m 和 j 的各种值，确定 $\boldsymbol{Y}_{m,t}$ 和 $\boldsymbol{Y}m,t-j-1$ 之间零值典型相关的结构可以用来标识 ARMA 模型的阶数 p 和 q, 因此，典型相关在模型识别中是有用的. 我们注意到典型相关函数的如下特殊情形。首先，当 $m=0$ 时，简化为去检验 z_t 和 z_{t-j-1} 之间的自相关 ρ_{j+1}, 对于一个 MA(q) 过程，$j\geq q$ 时这些值皆为零。其次，当 $j=0$ 的情形，我们就是去检验在给定 $z_{t-1},\cdots,z_{t-m}$ 时 z_t 和 z_{t-m} 之间的偏相关 $\phi_{m+1,m+1}$. $m\geq p$ 时，对于 AR(p) 过程，它们皆为零. 因此，典型相关分析可以看作是对于过程自相关和偏相关函数分析的推广.

在实际中，从 (6.2.4) 出发，引导人们去考虑样本典型相关函数 $\hat{\rho}(i)$，它是由下面矩阵的特征值所确定的：

$$(\sum_t \boldsymbol{Y}_{m,t}\boldsymbol{Y}'_{m,t})^{-1}(\sum_t \boldsymbol{Y}_{m,t}\boldsymbol{Y}'_{m,t-j-1})(\sum_t \boldsymbol{Y}_{m,t-j-1}\boldsymbol{Y}'_{m,t-j-1})^{-1} \cdot (\sum_t \boldsymbol{Y}_{m,t-j-1}\boldsymbol{Y}'_{m,t}) \qquad (6.2.6)$$

这里， $j=0,1,\cdots$ 及 $m=0,1,\cdots$ 对应于各种滞后。 Tsay 和 Tiao[193] 根据 (6.2.6) 的最小特征值 $\hat{\lambda}(m,j)$(样本典型相关的平方)，采用了一种 χ^2 检验统计量的方法。他们提出了统计量 $c(m,j)=-(n-m-j)\ln[1-\hat{\lambda}(m,j)/d(m,j)]$，其中 $d(m,j)=1+2\sum_{i=1}^{j}r_i^2(w'),j>0,r_i(w')$ 表示 $w'_t=z_t-\hat{\phi}_1^{(j)}z_{t-1}-\cdots-\phi_m^{(j)}z_{t-m}$ 的滞后为 i 的样本自相关，这里 $\hat{\phi}_i^{(j)}$ 是从对应于 $\hat{\lambda}(m,j)$ 的特征向量所得 ϕ_i 的估计值 [见方程 (6.2.5)]。他们证明了当 $m=p,j\geq q$，或 $m\geq p,j=q$ 时， $c(m,j)$ 有近似的 χ_1^2 分布。该统计量用来检验在理论上是否存在零值典型相关，从而如果样本统计量对于某些 p,q 值当 $m\geq p$ 和 $j\geq q$ 时的 χ_1^2 分布都呈现不显著的态势，那么，有理由将模型识别为 ARMA(p,q)，其中 (p,q) 是使上述态势成立的最小值。这种方法的另一特点是：Tsay 和 Tiao 还确定了该方法对非平稳 ARIMA 模型 $\varphi(B)z_t=\theta(B)a_t$ 也是有效的，因为广义 AR 算子 $\phi(B)$ 的总阶数可由该方法来确定，而无须在一开始就要决定对原序列 z_t 的差分。

Akaike[4],Cooper 和 Wood[72] 也提出了用典型相关方法进行 ARMA 建模。他们的方法就是在当前和过去值向量 $\boldsymbol{P}_t\equiv\boldsymbol{Y}_{m,t}=(\tilde{z}_t,\tilde{z}_{t-1},\cdots,\tilde{z}_{t-m})'$ 与未来值向量 $\boldsymbol{F}=(\tilde{z}_{t+1},\tilde{z}_{t+2},\cdots)'$ 之间进行典型相关分析。为构造当前过去值向量 $\boldsymbol{P}_t$ 所使用的有限滞后 m，我们可以通过采用例如 AIC 定阶准则来确定，在本节稍后将要对准则函数加以讨论，它用于拟合各种阶数的 AR 模型。典型相关分析是顺序进行的，每次增加元素到 $\boldsymbol{F}_{t+1}$ 中 (从 $\boldsymbol{F}^*_{t+1}=$

($\widetilde{z}_{t+1}$) 开始), 直到确定出 $\boldsymbol{P}_t$ 和 $\boldsymbol{F}^*_{t+1}$ 之间的第一个零值典型相关。Akaike[4] 用一种称为 DIC 的 AIC- 类准则来判断是否最小样本典型相关能够被视为零, 而 Cooper 和 Wood[72] 用一种传统的 χ^2 统计量方法来评价最小典型相关的显著性。当然, 正如 Tsay[191] 所指出的, 在滑动平均分量存在的情况下, 该统计量需要进行修正才是有效的。

在上述过程的某个给定阶段, 当 $\boldsymbol{P}_t$ 和 $\boldsymbol{F}^*_{t+1}$ 之间的典型相关第一次被判断为零, 且 $\widetilde{z}_{t+K+1}$ 是 $\boldsymbol{F}^*_{t+1}$ 中包含的最新近变量, 那么可以识别出用 $\boldsymbol{F}^*_{t+1}$ 的剩余元素表示的、$\widetilde{z}_{t+k+1}$ 的一个线性组合, 它与过去值是不相关的。特别地, 未来值向量 $\boldsymbol{F}^*_{t+1}$ 中元素的线性组合 $\widetilde{z}_{t+K+1}-\sum\limits_{j=1}^{K}\phi_j\widetilde{z}_{t+K+1-j}$ (在理论上) 被确定为与过去值 $\boldsymbol{P}_t$ 是不相关的。因此, 这种典型相关分析方法决定了过程的预报值 $\hat{z}_t(K+1)$ 满足

$$\hat{z}_t(K+1)-\sum_{j=1}^{K}\phi_j\hat{z}_t(K+1-j)=\theta_0$$

通过参考节 5.3 的 (5.3.2) 式, 对于一个平稳过程, 这意味着识别出了 $K=\max\{p,q\}$ 的 ARMA 模型。

正如在 Tsay 和 Tiao 的注记中可以看到的, Akaike,Cooper 和 Wood 的方法表示的是对 $n=1,2,\cdots$ 各个值所做的 $\boldsymbol{Y}_{m,t}$ 和 $\boldsymbol{Y}_{n-1,t+n}$ 之间的典型相关。而 Tsay 和 Tiao 的方法是对 $m=0,1,\cdots$ 和 $j=0,1,\cdots$ 各种组合考虑 $\boldsymbol{Y}_{m,t}$ 和 $\boldsymbol{Y}_{m,t-j-1}$ 之间的典型相关, 因此这是更为一般的方法, 且原则上能够分别给出模型 AR 和 MA 各自的阶数 p 和 q, 而不仅是这两个值中的最大者。在实际中, 当使用 Akaike,Cooper 和 Wood 的方法时, 关于 p,q 各自阶数更详细的信息可以在模型 ARMA(K,K) 参数极大似然估计的阶段来确定。

模型选择准则的使用　模型选择的另一种方法是使用信息准则, 诸如 Akaike[3] 提出的 AIC 或 Schwarz[176] 的 BIC 准则。当

使用该方法时，在可能的 ARMA 模型范围内用第 7 章中讨论的极大似然方法进行估计，对于每个模型，例如 AIC 准则值 (用样本大小 n 标准化) 由下式给出

$$\mathrm{AIC}_{p,q} = \frac{-2\ln(\text{极大似然值}) + 2r}{n} \simeq \ln(\hat{\sigma}_a^2) + r\frac{2}{n} + \text{常数}$$

相关的 BIC 由

$$\mathrm{BIC}_{p,q} = \ln(\hat{\sigma}_a^2) + r\frac{\ln(n)}{n}$$

计算，其中 $\hat{\sigma}_a^2$ 记 σ_a^2 的极大似然估计，而 $r = p + q + 1$ 表示模型中待估计参数的个数，模型中还包含一个常数项。在上述准则中，第一项主要对应于对数极大似然值的负 $2/n$ 倍，而第二项是对于在模型中包含附加参数的"惩罚因子"。在信息准则方法中，应选取产生最小准则值的模型，不同模型的 AIC,BIC 值要相互比较，以作为模型选择的基础。因此，与 AIC 准则相比， BIC 准则对于待估计模型参数数目赋予了较大的"惩罚"，所以，利用最小 BIC 进行模型选择时所选定的模型，其参数数目往往不多于在 AIC 下选出的模型。

这种方法的明显不利之处是：若干个模型都必须用极大似然方法去估计，在计算机上是很花费时间和精力的。正由于此，Hannan 和 Rissanen[103] 提出了下面的模型选择方法。该方法的第一阶段，通过用一个 m^*(足够高) 阶 AR 模型去近似未知的 ARMA 模型，以此得到随机冲击序列 a_t 的估计。用来逼近的 AR 模型的阶数 m^* 可以通过使用上述 AIC 模型选择准则来获得。由被选择的 $AR(m^*)$ 模型得到残差 $\tilde{a}_t = \tilde{w}_t - \sum_{j=1}^{m^*}\hat{\phi}_{m^*j}\tilde{w}_{t-j}$ 。该方法的第二阶段，是对各种 p,q 值做 $\tilde{w}_t$ 关于 $\tilde{w}_{t-1},\cdots,\tilde{w}_{t-p}$ 和 $\tilde{a}_{t-1},\cdots,\tilde{a}_{t-q}$ 的回归。即用普通最小二乘回归估计如下形式的近似模型

$$\tilde{w}_t = \sum_{j=1}^{p}\phi_j\tilde{w}_{t-j} - \sum_{j=1}^{q}\theta_j\tilde{a}_{t-j} + a_t \tag{6.2.7}$$

用 $\hat{\sigma}_{p,q}^2$ 记估计的误差方差 (未对自由度修正) 。因此通过应用上述 BIC 准则， ARMA 模型的阶 (p,q) 被选作使 $\ln(\hat{\sigma}_{p,q}^2)+(p+q)\ln(n)/n$ 达到最小。这种方法的好处在于：避免了在大范围内对许多可能的 ARMA 模型进行极大似然估计的计算。上述方法及模型选择中信息准则的使用是很有用的，但是这些方法应该看作是在模型选择过程中帮助抉择的补充。在实际中，不应用此代替对序列自相关、偏相关函数估计值特征的详细考察，也不应代替为鉴别模型是否失当而对残差 $\hat{a}_t$ 进行的严密考察，所有这些具体的考察都应作为主要部分贯穿于整个模型选择过程。

6.3 参数的初估计

6.3.1 由自协方差函数所得估计的唯一性

尽管一个给定的 ARMA 模型有唯一的协方差结构，其反命题却是不成立的。初看上去这会排除自协方差的估计值作为模型识别工具的作用。我们将在后面的 6.4 节证明，自协方差函数的估计值是可以用于模型识别的。其原因在于：尽管存在多个 ARMA 模型具有相同的自协方差函数，但只存在一个 ARMA 模型，它可以仅用过去的历史值来表示 $\nabla^d z_t = w_t$ 的当前值，并具有平稳可逆形式。

6.3.2 滑动平均过程的初估计

前文已经证明 (见 3.3.4),MA 过程的前 q 个自相关函数是非零的，且可以用模型的参数来表示为

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2} \quad (k=1,2,\cdots,q) \tag{6.3.1}$$

表达式 (6.3.1) 给出了用 $\theta_1,\theta_2,\cdots,\theta_q$ 表示 $\rho_1,\rho_2,\cdots,\rho_q$ 的 q 个未知量的 q 个方程。 θ 的初估计可以通过在 (6.3.1) 中以估计值 r_k

代替 ρ_k，并求解非线性方程得出。而 σ_a^2 的初估计可由下式求得

$$\gamma_0 = \sigma_a^2(1 + \theta_1^2 + \cdots + \theta_q^2)$$

上式中 θ 可用它的初估计代替，而 $\gamma_0 = \sigma_w^2$ 用其估计值代替。

(0, d, 1) 过程的初估计　第五部分的表 A 把 ρ_1 和 θ_1 联系起来，通过用 $r_1(w)$ 代替 ρ_1，便可将该表用于为 $(0, d, 1)$ 过程 $w_t = (1 - \theta_1 B)a_t$ 提供初估计，这里 $w_t = \nabla^d z_t$ 。

(0, d, 2) 过程的初估计　第五部分的图 C 把 ρ_1, ρ_2 和 θ_1, θ_2 联系起来，通过用 $r_1(w), r_2(w)$ 代替 ρ_1, ρ_2 该表可用于为任意 $(0, d, 2)$ 过程提供初估计。

在使用该方法得到初估计时，以下两点必须牢记：

1. 自协方差是 w 联合分布的二阶矩，而参数估计值是令样本矩与它们的理论值相等得到。众所周知，矩估计方法未必有效。可以证明，在一些特定的例子中，它是缺乏效率的。然而，由此得出的粗略估计对于得到充分有效的估计是有用的。因为对于大多数有效的估计方法来说，粗略估计给出了“在参数空间中向何处寻找”的大致意向。

2. 通常，等式 (6.3.1) 是用令矩值相等的方法求得，这就可能出现多重解。例如 $q = 1$ 时，

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2} \tag{6.3.2}$$

因此

$$\theta_1 = -\frac{1}{2\rho_1} + [\frac{1}{(2\rho_1)^2} - 1]^{1/2}$$

和

$$\theta_1 = -\frac{1}{2\rho_1} - [\frac{1}{(2\rho_1)^2} - 1]^{1/2} \tag{6.3.3}$$

表 6.5 序列 A,B 和 D 参数的初估计

序列	r_1	$\hat{\theta}_1$	$\hat{\lambda}_0 = 1 - \hat{\theta}_1$
A	-0.41	0.5	0.5
B	0.09	-0.1	1.1
C	-0.05	0.1	0.9

都是可能的解。这样，从表 6.2 可知，序列 A 一次差分后的一步滞后自相关约为 -0.4. 代入 (6.3.3) 可得 $\theta \simeq 0.5$ 和 $\theta_1' \simeq 2.0$ 。但是 $\theta_1 \simeq 0.5$ 是在可逆域 $-1 < \theta_1 < 1$ 内的唯一待选值。事实上，在 6.4.1 节中证明了：多重解中只有一个满足可逆性条件，这一点始终是正确的。

例：序列 A,B 和 D 在表 6.4 中都被识别为可能是 IMA(0,1,1) 过程。在 4.3.1 节中已看到，这个模型可以被写为另一种形式：

$$\nabla z_t = (1 - \theta_1 B)a_t$$

$$\nabla z_t = \lambda_0 a_{t-1} + \nabla a_t \qquad (\lambda_0 = 1 - \theta_1)$$

$$z_t = \lambda_0 \sum_{j=1}^{\infty}(1 - \lambda_0)^{j-1} z_{t-j} + a_t$$

用第五部分的表 A, 可得到表 6.5 列出的参数近似估计值。

序列 C 在表 6.4 中被试探识别为 IMA(0,2,2) 过程

$$\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$$

或等价地

$$\nabla^2 z_t = (\lambda_0 \nabla + \lambda_1)a_{t-1} + \nabla^2 a_t$$

由于在表 6.2 中已给出 $\nabla^2 z_t$ 的前两阶自相关近似为零，那么，利用第五部分图 C, $\hat{\theta}_1 = 0, \hat{\theta}_2 = 0$, 从而有 $\hat{\lambda}_0 = 1 + \hat{\theta}_2 = 1$ 及 $\hat{\lambda}_1 = 1 - \hat{\theta}_1 - \hat{\theta}_2 = 1$, 在这基础上，序列 C 表示为

$$\nabla^2 z_t = a_t \tag{6.3.4}$$

这意味着二阶差分后的序列 $\nabla^2 z_t$ 近似为一个随机序列。

6.3.3 自回归过程的初估计

对于假设其阶数为 1 或 2 的 AR 过程来说，ϕ_1 和 ϕ_2 的初估计可以通过用估计值 r_j 代替表 6.1 公式中的理论相关函数 ρ_j 而计算出来，该式是从 Yule-Walker 方程 (3.2.6) 得到的。特别地，对于 AR(1) 过程，$\hat{\phi}_{11}=r_1$, 而对于 AR(2) 过程有

$$\hat{\phi}_{21}=\frac{r_1(1-r_2)}{1-r_1^2}, \qquad \hat{\phi}_{22}=\frac{r_2-r_1^2}{1-r_1^2} \tag{6.3.5}$$

这里，ϕ_{pj} 表示 p 阶过程中的第 j 个自回归参数。在 (3.2.7) 中用 r_j 代替 ρ_j, 可得由 Yule-Walker 方程高阶形式给出的相应公式：

$$\hat{\boldsymbol{\phi}}=\boldsymbol{R}_p^{-1}\boldsymbol{r}_p \tag{6.3.6}$$

这里，$\boldsymbol{R}_p$ 是直到 $p-1$ 阶的 $p\times p$ 维相关阵的估计，$\boldsymbol{r}_p'$ 是向量 $(r_1,r_2,\cdots,r_p)$ 。例如，若 $p=3$ 则 (6.3.6) 为

$$\begin{bmatrix}\hat{\phi}_{31}\\ \hat{\phi}_{32}\\ \hat{\phi}_{33}\end{bmatrix}=\begin{bmatrix}1 & r_1 & r_2\\ r_1 & 1 & r_1\\ r_2 & r_1 & 1\end{bmatrix}^{-1}\begin{bmatrix}r_1\\ r_2\\ r_3\end{bmatrix} \tag{6.3.7}$$

关于从 AR$(p-1)$ 过程得到 AR(p) 过程参数估计值的递推算法，在附录 A3.2 中给出。

在第 7 章中将证明，与 MA 过程的情况相对照，由 (6.3.6) 所得到的自回归参数渐近于充分有效的极大似然估计。

例：序列 E 的未经差分形式类似于二阶自回归过程[①]

$$(1-\phi_1B-\phi_2B^2)\tilde{z}_t=a_t$$

[①] 太阳黑子是许多研究的主题，如见 Schuster[175],Yule[212] 和 Moran [146]) . 这个序列几乎肯定不能由二阶自回归过程恰当地拟合. 一个与实际运作机制相联系的模型当然最能令人满意. 最近一些未出版的著作建议，从经验角度来看，如果首先对 z_t 实施某种适当的变换，再用一个二阶自回归过程将会给出更好的拟合. 若在 AR 模型中包含高阶项 (9 步滞后), 也会改进拟合. 其它的可能性包括使用非线性时间序列模型，比如双线性或门限自回归模型，正如 Subba Rao 和 Gabr[184],Tong 和 Lim[189] 和 Tong[188] 所从事的研究.

将从表 6.2 得到的 $r_1 = 0.81$ 和 $r_2 = 0.43$ 代入 (6.3.5) 中，我们有 $\hat{\phi}_1 = 1.32$ 及 $\hat{\phi}_2 = -0.63$ 。

作为第二个例子，重新考虑序列 C, 它被识别为 (1,1,0), 或可能为 (0,2,2) 。前一种可能将给出

$$(1-\phi_1 B)\nabla z_t = a_t$$

由于 ∇z 的 r_1 是 0.81, 所以 $\hat{\phi}_1 = 0.8$ 。

这个例子是特别有意义的，因为它清楚地显示出由该序列识别出的两个模型之间的密切联系. 在序列为 (0,2,2) 阶的假设下，我们给出模型

$$(1-B)(1-B)z_t = a_t \tag{6.3.8}$$

另一个模型为

$$(1-0.8B)(1-B)z_t = a_t \tag{6.3.9}$$

二者是非常相似的。

6.3.4 自回归滑动平均混合模型的初估计

我们经常发现，无论对于原始的或经过适当差分后的序列，对 $w_t = \nabla^d z_t$ 最经济的描述是采用混合的 ARMA 过程

$$\phi(B)w_t = \theta(B)a_t$$

在 6.2.1 节中提到，如果自相关和偏相关函数都呈现“拖尾”性，则意味着一个混合过程. 另一个有助于识别混合过程的事实为 [见 (3.4.3)]: 混合过程 $q-p$ 步滞后的理论自相关函数与纯自回归过程 $\phi(B)w_t = a_t$ 的自相关函数相似。特别地，如果 d 阶差分后的自相关函数除了第一个特殊值 r_1 之外，呈现指数下降，那么这是一个 $(1,d,1)$ 过程，即

$$(1-\phi_1 B)w_t = (1-\theta_1 B)a_t \tag{6.3.10}$$

这里， $w_t = \nabla^d z_t$ 。

用估计值 $r_1(w)$ 和 $r_2(w)$ 代替表达式 (3.4.8) 和表 6.1 中所示的 ρ_1 和 ρ_2, 可以得到过程 (6.3.10) 参数的近似值，因此有

$$r_1 = \frac{(1-\hat{\theta}_1\hat{\phi}_1)(\hat{\phi}_1-\hat{\theta}_1)}{1+\hat{\theta}_1^2-2\hat{\phi}_1\hat{\theta}_1}$$

$$r_2 = r_1\hat{\phi}_1$$

第五部分的表 C 把 ρ_1, ρ_2 与 ϕ_1, θ_1 联系起来，且可以用于为任何 $(1, d, 1)$ 过程的参数提供初估计。

例如根据图 6.2, 序列 A 被识别为 (0,1,1) 阶的过程， θ_1 约为 0.5. 观察 z_t 而不是 $w_t = \nabla z_t$ 的自相关函数，我们看到，从 r_1 开始自相关函数虽然衰减较慢，但粗略地为指数衰减。这样对于序列 A 的另一种识别是：它由一个 (1,0,1) 阶的平稳过程所产生。自相关函数估计值和相应参数初估计为

$$r_1 = 0.57 \qquad r_2 = 0.50 \qquad \hat{\phi}_1 \simeq 0.87 \qquad \hat{\phi}_2 \simeq 0.48$$

而在表 6.5 中给出最初识别的模型是 (0,1,1) 阶的：

$$(1-0.9B)\tilde{z}_t = (1-0.5B)a_t$$

我们又一次看到，两个待选的模型几乎是一样的。有关自回归滑动平均混合模型获得参数初估计的更一般方法在附录 A6.2 中给出。

自回归与滑动平均算子之间的补偿　上述识别出的两个待选模型看起来非常相似，这是因为对于一个混合模型，其自回归算子中的微小变化可以通过滑动平均算子中相应的变化来近似地补偿。特别地，如果我们有一个模型

$$[1-(1-\delta)B]\tilde{z}_t = (1-\theta B)a_t$$

其中 δ 是一个小的正数，我们可以写成

$$\begin{aligned}(1-B)\tilde{z}_t &= [1-(1-\delta)B]^{-1}(1-B)(1-\theta B)a_t\\ &= \{1-\delta B[1+(1-\delta)B+(1-\delta)^2B^2+\cdots]\}(1-\theta B)a_t\\ &= [1-(\theta+\delta)B]a_t + (\text{ 含}a_{t-2}, a_{t-3}, \cdots \text{ 和}\delta\text{ 幂次的项})\end{aligned}$$

6.3.5 在不确定场合对平稳和非平稳模型的抉择

从表 6.4 中看到, 在对序列 A,C 和 D 识别模型时表现出明显的不确定性, 这当然比实际的情形更显著。只要 $\phi(B) = 0$ 的根接近于 1 就会出现这种情形. $\phi(B)$ 是否包含接近于 1 的根, 或与此相应地, 是否应该再作一次差分, 在这种场合就变得不重要了。利用第 7 章中讨论的估计方法可以进行更精确的评价, 但是必须牢记以下几点:

1. 时间序列的长度必然是有限的, 由此出发要证明一个自回归算子的零点严格等于 1, 这是不可能的。

2. 当然不会存在从平稳到非平稳特性的突然转变, 这一点可以通过考虑如下简单混合模型的特点来理解:

$$(1 - \phi_1 B)(z_t - \mu) = (1 - \theta_1 B)a_t$$

由这样一个模型所产生的序列, 随着 ϕ_1 趋于 1 其表现越来越不平稳. 例如, 一个 $\phi_1 = 0.99$ 的序列可能游离于其均值 μ, 并且在很长时期内不返回。这就象是随着 ϕ_1 接近于 1, 均值 μ 施加在序列上的吸引力越来越减弱。最后, 当 ϕ_1 等于 1 时, 序列的性质就与 μ 完全独立了。

对于不确定的情形, 选用非平稳模型来代替平稳模型 (即当 ϕ_1 的估计值接近于 1 时, 就按等于 1 处理) 可能更为有利, 在预测和控制问题中更是如此。在 ϕ_1 接近于 1 时, 我们并不知道序列的期望是否有意义, 因此, 采用一个不包含期望 μ 的非平稳模型是有利的。如果我们用了一个这样的模型, 那么对未来行为的预测就不会以任何方式依赖于期望的估计值, 而该估计值是由过去值计算出的, 它与未来值的水平并无关联。

6.3.6 对 ARIMA 模型单位根更正规的检验

正如前文所指出的, 到底是否需要进行差分运算, 若采用非正式的方法, 可以建立在检验时间序列图和样本自相关函数特征的基础之上 (例如, r_k 不是足够快地衰减) 。这还可以根据第 7 章

所讨论的模型参数的有效估计来进一步评价. 但是, 必须牢记, 参数估计的分布理论在平稳与非平稳模型之间差别很大. 由此引起人们对模型中有关差分算子 (或 AR 算子中的单位根) 的合理性采用更正式推断方法的兴趣. 这涉及在单位根非平稳时间序列模型中有关的估计、检验以及参数估计值的分布理论.

首先考虑简单的 AR(1) 模型 $z_t=\phi z_{t-1}+a_t, t=1,2,\cdots,n, z_0=0$. 考虑对随机游动模型的检验, 即 $\phi=1$. ϕ 的最小二乘估计由下式给出

$$\hat{\phi}=\frac{\sum_{t=2}^{n} z_{t-1} z_t}{\sum_{t=2}^{n} z_{t-1}^2}=\phi+\frac{\sum_{t=2}^{n} z_{t-1} a_t}{\sum_{t=2}^{n} z_{t-1}^2}$$

在平稳情形有 $|\phi|<1$, 在第 7 章中曾指出 $n^{1/2}(\hat{\phi}-\phi)$ 渐近为标准正态分布, 期望为 0, 方差为 $(1-\phi^2)$. 然而, 当 $\phi=1$ 时有求和形式: $z_t=\sum_{j=0}^{t-1} a_{t-j}+z_0$, 于是可表示为

$$n(\hat{\phi}-1)=n^{-1}\sum_{t=2}^{n} a_{t-1} z_t / n^{-2}\sum_{t=2}^{n} z_{t-1}^2=O_p(1)$$

当 $n\to\infty$ 时, 该式为依概率有界, 分子和分母都具有非退化和非正态的极限分布. 因此, 在非平稳的情形, 随着样本数 n 的增大, 估计量 $\hat{\phi}$ 比平稳情形以更快的速度趋于真值 $\phi=1$.

Dickey 和 Fuller[82] 给出了 $n(\hat{\phi}-1)$ 极限分布的表达式:

$$n(\hat{\phi}-1)\xrightarrow{D}\frac{\frac{1}{2}(\Lambda^2-1)}{\Gamma} \tag{6.3.11}$$

这里, $(\Gamma,\Lambda)=(\sum_{i=1}^{\infty}\gamma_i^2 Z_i^2, \sum_{i=1}^{\infty} 2^{1/2}\gamma_i Z_i)$, 其中 $\gamma_i=2(-1)^{i+1}/[(2i-1)\pi]$, 且 Z_i 是独立并具有相同 $N(0,1)$ 分布的随机变量. 该分布的另一等价形式为 (见 Chan 和 Wei[68])

$$n(\hat{\phi}-1)\xrightarrow{D}\frac{\int_0^1 B(u)dB(u)}{\int_0^1 B(u)^2 du}=\frac{\frac{1}{2}(B(1)^2-1)}{\int_0^1 B(u)^2 du} \tag{6.3.12}$$

这里，$B(u)$ 是 $[0,1]$ 上的 (连续参数) 标准布朗运动过程。Fuller 给出了 $n(\hat{\phi}-1)$ 极限分布百分位表。另外，考虑"学生"统计量

$$\hat{\tau}=(\hat{\phi}-1)/\{s_a(\sum_{t=2}^{n}z_{t-1}^2)^{-1/2}\} \tag{6.3.13}$$

其中 $s_a^2=(n-2)^{-1}(\sum_{t=2}^{n}z_t^2-\hat{\phi}\sum_{t=2}^{n}z_{t-1}z_t)$ 为残差均方。统计量 $\hat{\tau}$ 的极限分布可以导出。在 [93] 中有在 $\phi=1$ 之下该分布的百分位表，可用于有关随机游动的假设：$\phi=1$。当 $\hat{\tau}$"负得太多"时，则拒绝 $\phi=1$。

对于高阶 (广义)AR 过程 $\phi(B)z_t=a_t$，考虑 $\varphi(B)=\phi(B)(1-B)$ 的情形，且 $\phi(B)=1-\sum_{j=1}^{p}\phi_jB^j$ 是平稳 p 阶 AR 算子。于是有

$$\varphi(B)z_t=\phi(B)(1-B)z_t=z_t-z_{t-1}-\sum_{j=1}^{p}\phi_j(z_{t-j}-z_{t-j-1})$$

对 $\varphi(B)$ 作单位根检验等价于在模型 $z_t=\rho z_{t-1}+\sum_{j=1}^{p}\phi_j(z_{t-j}-z_{t-j-1})+a_t$，或在模型

$$z_t-z_{t-1}=(\rho-1)z_{t-1}+\sum_{j=1}^{p}\phi_j(z_{t-j}-z_{t-j-1})+a_t \tag{6.3.14}$$

中检验 $\rho-1=0$。事实上，任何 (广义)AR$(p+1)$ 模型 $z_t=\sum_{j=1}^{p+1}\varphi_jz_{t-j}+a_t$ 可以表示为等价形式：$w_t=(\rho-1)z_{t-1}+\sum_{j=1}^{p}\phi_jw_{t-j}+a_t$，其中 $w_t=z_t-z_{t-1},\rho-1=-\varphi(1)=\sum_{j=1}^{p+1}\varphi_j-1$，且 $\phi_j-\sum_{i=j+1}^{p+1}\varphi_j$。由此可知，在 AR 算子 $\varphi(B)$ 中存在单位根等价于 $\rho=\sum_{j=1}^{p+1}\varphi_j=1$。因此，基于该模型最后的表示，用 $(\hat{\rho}-1,\hat{\phi}_1,\cdots,\hat{\phi}_p)$ 表示 (6.3.14) 中模型的普通最小二乘估计，它们是用 $w_t=z_t-z_{t-1}$ 对 $z_{t-1},w_{t-1},\cdots,w_{t-p}$ 作回归而得到的，这里，$w_t=z_t-z_{t-1}$。那么，在 $\rho=1$ 的单位根模型和 $\phi(B)$ 平稳的假设下，已经证明 ([93] 定理 8.5.1 和系 8.5.1) 了：$(\hat{\rho}-1)/s_a\sum_{t=p+2}^{n}z_{t-1}^2)^{-1/2}$ 与 (6.3.13) 中 AR(1)

情形下的统计量 $\hat{\tau}$ 具有相同的极限分布。而 $(n-p-1)(\hat{\rho}-1)c$, 这里 $c=\sum_{j=0}^{\infty}\psi_j$, 及 $\psi(B)=\phi^{-1}(B)$ 与 AR(1) 情形下的统计量 $n(\hat{\phi}-1)$ 具有相同的渐近分布. 另外, 还可以用 $(\hat{\rho}-1)$ 除以由最小二乘回归得出的“普通估计标准差”, 由此构造出的统计量记为 $\hat{\tau}$, 统计量 $(\hat{\rho}-1)/s_a(\sum_{t=p+2}^{n} z_{t-1}^2)^{-1/2}$ 与 $\hat{\tau}$ 渐近相等。因此, 它和 AR(1) 情形下的 $\hat{\tau}$ 有相同的极限分布。

如上所述对统计量 $\hat{\tau}$ 的检验可以用来检验关于 AR$(p+1)$ 模型中 $\rho=1$ 的假设 (即在 AR$(p+1)$ 模型 $\varphi(B)z_t=a_t$ 中的单位根检验). 而且进一步证明了 [93 定理 8.5.1], 在模型中对“平稳算子”$\phi(B)$ 的参数得到最小二乘估计 $(\hat{\phi}_1,\cdots,\hat{\phi}_p)$, 及对平稳差分序列 $w_t=z_t-z_{t-1}$ 关于 $w_{t-1},\cdots,w_{t-p}$ 作回归得出的最小二乘估计 $(\widetilde{\phi}_1,\cdots,\widetilde{\phi}_p)$, 此二者具有相同的标准渐近分布 (对于平稳 AR 模型的估计将在第 7 章中讨论). 我们也注意到, 上述结果也可以推广到包括一个常数项的最小二乘估计的情形, 类似于 $\hat{\tau}$ 的统计量用 $\hat{\tau}_\mu$ 来记, 当然, $\hat{\tau}_\mu$ 的极限分布是在 $\rho=1$ 的假设之下当模型中常数项 $\theta_0=(1-\varphi_1-\cdots-\varphi_{p+1})\mu=(1-\rho)\mu$ 的“真值”等于 0 而推出的。例如, 在 AR(1) 模型 $z_t=\phi z_{t-1}+\theta_0+a_t$ 中, 我们得到最小二乘估计:

$$\hat{\phi}_\mu=\frac{\sum_{t=2}^{n}(z_{t-1}-\overline{z}_{(1)})(z_t-\overline{z}_{(0)})}{\sum_{t=2}^{n}(z_{t-1}-\overline{z}_{(1)})^2}, \quad \text{其中 } \overline{z}_{(i)}=(n-1)^{-1}\sum_{t=2}^{n}z_{t-i}, i=0,1$$

类似于 (6.3.12) 的极限分布的表达式由下式给出

$$n(\hat{\phi}_\mu-1)\xrightarrow{D}\frac{\int_0^1 B(u)dB(u)-\xi B(1)}{\int_0^1 B(u)^2du-\xi^2}$$

这里, $\xi=\int_0^1 B(u)du$, 并且假设当 $\phi=1$ 时 $\theta_0=(1-\phi)\mu=0$。在 AR(1) 过程中对 $\phi=1$ 的相应检验统计量为

$$\hat{\tau}_\mu = \frac{\hat{\phi}_\mu - 1}{s_a[\sum_{t=2}^{n}(z_{t-1} - \overline{z}_{(1)})^2]^{-1/2}} \tag{6.3.15}$$

$\phi = 1$ 时 $\hat{\tau}_\mu$ 分布的百分位表可见 [93] 。注意，在 $\phi = 1$ 的条件下，由于在截断随机冲击或求和形式中有 $z_t = \sum_{j=0}^{t-1} a_{t-j} + z_0$, 所以 $z_t - \overline{z}_{(0)}$ 和 $z_{t-1} - \overline{z}_{(1)}$ 项中不包括初始值 z_0, 从而可看出 $\hat{\phi}_\mu$ 最小二乘估极的分布理论不依赖于有关 z_0 的假设。

上述假验方法及一些类似方法被推广到混合 ARIMA$(p,1,q)$ 模型的单位根检验 (例如，可见 Said 和 Dickey[172] 和 Solo[181])，以及用于对高阶差分作检验 (例如见 Dickey 和 Pantula[83]) 。另外，Dickey,Bell 和 Miller[81] 给出了有关单变量 ARIMA 模型单位根检验各种结果的评论。对于作差分没有把握的情形，这些正规的检验方法可以用来补充其它的信息。并且，只有当发现了充分的证据能否定单位根假设的情形下，选择平稳模型才是明智的。

例 ： 为了说明问题，我们对序列 C 考虑两个试探性检验的模型： ARIMA(1,1,0) 和 ARIMA(0,2,0) 。因为在 ARIMA(0,2,0) 模型中要作二阶差分，是否果真需要这样做是存在疑问的，况且另一个模型在一阶差分后是平稳 AR(1) 模型，因此应对此做更正式地研究。一阶差分的 AR(1) 模型 $\nabla z_t = \phi \nabla z_{t-1} + a_t$ 可以改写为 $\nabla^2 z_t = (\phi - 1)\nabla z_{t-1} + a_t$, 用这种形式可得最小二乘回归估计 $\hat{\phi} - 1 = -0.187$, 且估计的标准差为 0.038,$\hat{\sigma}_a^2 = 0.018$ 。我们注意到，这意味着 $\hat{\phi} = 0.813$, 与表 6.7 相似。检验 $\phi = 1$ 的 "学生 " 统计量是 $\hat{\tau} = -4.87$, 它比 [93] 的表中的下百分位点 -2.58 的绝对值大了很多。因此这些结果不支持作二阶差分，而倾向于使用 ARIMA(1,1,0) 模型。

为了与第 7 章中所述更有效的估计方法作比较，考察一下在识别阶段可以提取出多少有关模型的补充信息是有意义的。我们已经说明了怎样对序列的恰当差分加以识别，以及如何得到 ARMA 模型参数 $(\hat{\phi}, \hat{\theta})$ 的初估计。为了完成这一工作，我们现在

来说明如何得到残差方差 σ_a^2 的初始估计以及恰当差分序列均值的近似标准差.

6.3.7 残差方差的初估计

用估计值 c_0 代替第 3 章所给方差表达式中的 γ_0, 可得到残差方差的初估计. 因此, 在 (3.2.8) 中作代换, 可以得到 AR 过程的 σ_a^2 的初估计

$$\hat{\sigma}_a^2 = c_0(1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p) \tag{6.3.16}$$

类似地, 从 (3.3.3) 可的 MA 过程 σ_a^2 的初估计为

$$\hat{\sigma}_a^2 = \frac{c_0}{1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2} \tag{6.3.17}$$

对混合过程而言, σ_a^2 的估计形式要更复杂一些, 但在附录 A6.2 中所述的是最简便的求取方法. 对于重要的 ARMA(1,1) 过程来说, 其形式为 (见 (3.4.7))

$$\hat{\sigma}_a^2 = \frac{1 - \hat{\phi}_1^2}{1 + \hat{\theta}_1^2 - 2\hat{\phi}_1\hat{\theta}_1} c_0 \tag{6.3.18}$$

例如, 考虑对序列 A 所识别的 (1,0,1) 模型, 由 $\hat{\phi}_1 = 0.87, \hat{\theta}_1 = 0.48$ 和 $c_0 = 0.1586$, 利用 (6.3.18) 我们得到该估计为

$$\hat{\sigma}_a^2 = \frac{1 - (0.87)^2}{1 + (0.48)^2 - 2(0.87)(0.48)} 0.1586 = 0.098$$

6.3.8 $\overline{w}$ 的近似标准差

对于一般的 ARIMA 模型 ,$w_t = \nabla^d z_t$ 的均值 μ_w 不一定为 0, 可将模型写为下面三种形式之一:

$$\phi(B)(w_t - \mu_w) = \theta(B) a_t \tag{6.3.19}$$

$$\phi(B) w_t = \theta_0 + \theta(B) a_t \tag{6.3.20}$$

$$\phi(B) w_t = \theta(B)(a_t + \xi) \tag{6.3.21}$$

其中

$$\mu_w = \frac{\theta_0}{1-\phi_1-\phi_2-\cdots-\phi_p} = \frac{(1-\theta_1-\theta_2-\cdots-\theta_q)\xi}{1-\phi_1-\phi_2-\cdots-\phi_p}$$

因此, 若 $1-\phi_1-\phi_2-\cdots-\phi_p \neq 0$ 且 $1-\theta_1-\theta_2-\cdots-\theta_q \neq 0, \mu_w = 0$, 蕴含 $\theta_0 = 0$ 及 $\xi = 0$ 。于是, 一般地, 当 $d = 0$ 时 μ_w 并不为 0. 然而, 考虑当 $d > 0$ 时与一般模型 (6.3.19) 有联系的最终预报函数, $\mu_w = 0$ 时, 该预报函数已经包含了一个自适应的 $d-1$ 阶的多项式分量. 允许 μ_w 不为 0, 其效果是在预报函数中引入了一个固定的 d 次多项式. 例如, 若 $d = 2, \mu_w$ 非零, 则预报函数 $\hat{z}_t(l)$ 包含了一个 l 的二次分量, 而这个二次项的系数是固定的, 且不是自适应于该序列. 由于当 $d > 0$ 时这种类型的模型通常是无法使用的, 故 $\mu_w = 0$ 的假设通常不会与数据相矛盾. 事实上, 正如我们已指出的, 除非有相反的证据, 我们通常仍假设 $\mu_w = 0$ 。

为此, 在建模的识别阶段, 关于 μ_w 是否非零需要有一种标识, 我们可以通过将 $\overline{w} = \sum w_t/n$ 与其近似标准差 (见 2.1.5) 做比较而得到. 当有 $n = N - d$ 个差分项时, 有

$$\sigma^2(\overline{w}) = n^{-1}\gamma_0 \sum_{-\infty}^{+\infty} \rho_j = n^{-1} \sum_{-\infty}^{+\infty} \gamma_j$$

即

$$\sigma^2(\overline{w}) = n^{-1}\gamma(1) \tag{6.3.22}$$

这里, $\gamma(B)$ 是 (3.1.10) 所定义的自协方差生成函数, $\gamma(1)$ 是将 $B = B^{-1} = 1$ 代入的值。

为了说明上述结果, 考虑 $(1, d, 0)$ 阶过程

$$(1-\phi B)(w_t - \mu_w) = a_t$$

其中 $w_t = \nabla^d z_t$ 。从 (3.1.11) 得到

$$\gamma(B) = \frac{\sigma_a^2}{(1-\phi B)(1-\phi F)}$$

于是

$$\sigma^2(\overline{w}) = n^{-1}(1-\phi)^{-2}\sigma_a^2$$

而 $\sigma_a^2 = \sigma_w^2(1-\phi^2)$, 因此有

$$\sigma^2(\overline{w}) = \frac{\sigma_w^2}{n}\frac{1-\phi^2}{(1-\phi)^2} = \frac{\sigma_w^2}{n}\frac{1+\phi}{1-\phi}$$

及

$$\sigma(\overline{w}) = \sigma_w\left[\frac{1+\phi}{n(1-\phi)}\right]^{1/2}$$

现在, 如 (2.1.9) 和 (2.1.10) 所定义的, σ_w^2 和 ϕ 可分别由 c_0 和 r_1 估计出来. 因此对于一个 $(1,d,0)$ 过程, 所需的标准差由下式给出

$$\hat{\sigma}(\overline{w}) = \left[\frac{c_0(1+r_1)}{n(1-r_1)}\right]^{1/2}$$

完全类似地可得到表 6.6 中给出的 $\hat{\sigma}(\overline{w})$ 的表达式

模型 A 到 F 的试探性识别　表 6.7 概括了对序列 A 到 F 试探性识别的模型, 并将参数的初值代入。这些模型可作为第 7 章中讨论的更有效估计方法的初步推测.

表 6.6　$\overline{w}$ 的近似标准差, 其中 $w_t = \nabla^d z_t$, 且 z_t 是 (p,d,q) 阶的 ARIMA 过程

$(1,d,0)$	$(0,d,1)$
$\left[\frac{c_0(1+r_1)}{n(1-r_1)}\right]^{1/2}$	$\left[\frac{c_0(1+2r_1)}{n}\right]^{1/2}$
(2,d,0)	(0,d,2)
$\left[\frac{c_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}\right]^{1/2}$	$\left[\frac{c_0(1+2r_1+2r_2)}{n}\right]^{1/2}$
$(1,d,1)$	
$\left[\frac{c_0}{n}\left(1+\frac{2r_1^2}{r_1-r_2}\right)\right]^{1/2}$	

表 6.7 对序列 A 到 F 识别模型的总结，其中参数用初估计代入

序列	差分阶数	$\overline{w} \pm \hat{\sigma}(\overline{w})^*$	$\hat{\sigma}_w^2 = c_0$	识别模型	$\hat{\sigma}_a^2$
A	0	17.06±0.10	0.1586	$z_t-0.87z_{t-1}=2.45+a_t-0.48a_{t-1}$	0.098
	或 1	0.002±0.011	0.1364	$\nabla z_t=a_t-0.53a_{t-1}$	0.107
B	1	−0.28±0.41	52.54	$\nabla z_t=a_t+0.09a_{t-1}$	52.2
C	1	−0.035±0.047	0.0532	$\nabla z_t-0.81z_{t-1}=a_t$	0.019
	或 2	−0.003±0.008	0.0198	$\nabla^2 z_t=a_t-0.09a_{t-1}-0.07a_{t-2}$	0.020
D	0	9.13±0.04	0.3620	$z_t-0.86z_{t-1}=1.32+a_t$	0.093
	或 1	0.004±0.017	0.0965	$\nabla z_t=a_t-0.05a_{t-1}$	0.096
E	0	46.9±5.4	1382.2	$z_t-1.32z_{t-1}+0.63z_{t-2}=14.9+a_t$	289
	或 0	46.9±5.4	1382.2	$z_t-1.37z_{t-1}+0.74z_{t-2}-0.08z_{t-1}=13.7+a_t$	287 287
F	0	51.5±1.1	139.80	$z_t+0.32_{t-1}-0.18z_{t-1}=58.3+a_t$	115

* 当 $d=0$ 时，用 z 代替 w .

6.4 模型的多重性

6.4.1 自回归滑动平均模型的多重性

在正态假设下，一个概率分布的一阶矩和二阶矩便包含了这个分布的全部信息。特别地，$w_t = \nabla^d z_t$ 的均值及自协方差函数的信息即可唯一地确定 w_t 的概率结构。尽管有多个线性模型都可能表示同一概率结构，然而，当我们适当地引入平稳性和可逆性限制后，可以实现模型的唯一性。

假设 w_t 有协方差生成函数 $\gamma(B)$, 且线性模型表示为

$$\phi(B)w_t = \theta(B)a_t \tag{6.4.1}$$

$\phi(B)$ 和 $\theta(B)$ 的零点都在单位圆外。于是该模型可表为

$$\prod_{i=1}^{p}(1 - G_iB)w_t = \prod_{j=1}^{q}(1 - H_jB)a_t \tag{6.4.2}$$

这里，G_i^{-1} 是 $\phi(B)=0$ 的根，H_j^{-1} 是 $\theta(B)=0$ 的根，且 G_i,H_j 都在单位圆内。利用 (3.1.11) 可知 w 的协方差生成函数是

$$\gamma(B)=\prod_{i=1}^{p}(1-G_iB)^{-1}(1-G_iF)^{-1}\prod_{j=1}^{q}(1-H_jB)(1-H_jF)\sigma_a^2$$

滑动平均模型的多重选择　由于

$$(1-H_j)(1-H_jF)=H_j^2(1-H_j^{-1}B)(1-H_j^{-1}F)$$

可知任何如下形式的随机模型

$$\prod_{i=1}^{p}(1-G_iB)w_t=\prod_{j=1}^{q}(1-H_j^{\pm 1}B)ka_t$$

只要适当选择常数 k，便具有相同的协方差生成函数。在上式中，对于复根来说，共轭对中两者的倒数都要取，这是可以理解的。因此，如果一个实根 H 在单位圆内，则 H^{-1} 就在单位圆外，或者，一对复根 H_1 和 H_2 在单位圆内，则 H_1^{-1} 和 H_2^{-1} 就在单位圆外。由此可知，只有一个形如 (6.4.2) 的平稳且可逆模型，它具有给定的自协方差函数。

反向表达式　现在，如果我们在 (6.4.2) 中用 $1-G_iF$ 代替 $1-G_iB$，或是用 $1-H_jF$ 代替 $1-H_jB$，$\gamma(B)$ 仍保持不变。因此所有的随机模型

$$\prod_{i=1}^{p}(1-G_iB^{\pm 1})w_t=\prod_{j=1}^{q}(1-H_jB^{\pm 1})a_t$$

都有相同的协方差结构。然而，含有算子 $B^{-1}=F$ 的表达式涉及到未来的 w 和 (或) a。因此，虽然存在 w_t 按未来的 w 和 a 展开的平稳、可逆的表达式，但是只存在一个表达式 (6.4.2)，它使 w_t 只与过去的历史值相联系。

较为特别而有实际意义的模型是在 (6.4.1) 中用 F 代替 B:

$$\phi(F)w_t = \theta(F)e_t$$

这里，e_t 是独立分布的随机变序列，其均值为 0, 方差 $\sigma_e^2 = \sigma_a^2$。这是一个平稳可逆的表达式，其中，w 完全用未来的 w 和 e 来表示。我们将这种形式称为过程的反向形式，或者简单地称为反向过程。

方程 (6.4.2) 并不是具有协方差生成函数 $\gamma(B)$ 的平稳可逆线性模型的最一般形式，例如，模型 (6.4.2) 可在两边乘以因式 $1 - QB$, 因此过程

$$(1 - QB)\prod_{i=1}^{p}(1 - G_iB)w_t = (1 - QB)\prod_{j=1}^{q}(1 - H_jB)a_t$$

具有和 (6.4.2) 相同的协方差结构. 这一事实不会对模型识别带来特别的困难，因为我们会自然地倾向于选择最简单的表达式. 然而，我们在第 7 章中会发现，在拟合过程时应注意算子因式分解的可能性.

最后，我们的结论是：一个平稳可逆的模型，若其当前值 w_t 仅用既往历史值来表示，且模型不包含公共因子，那么，该模型唯一地被协方差结构所确定.

正确地理解模型的多重性是很重要的，原因如下：

1. 我们再次重申前面的结论：协方差函数在逻辑上可用于识别一个线性平稳可逆模型，该模型用过去历史值来表示 w_t

2. 滑动平均参数的多重解可通过令矩值相等而求得，相应多重解的性质已经阐明.

3. 反向过程

$$\phi(F)w_t = \theta(F)e_t$$

是在线性模型中用 F 代替 B 而得到，该过程对于估计发生在第一个观测值取得之前的序列值是有用的.

下面我们更详细地讨论 (2) 和 (3)。

6.4.2 滑动平均参数的多重矩法求解

在通过使协方差相等去估计 MA 模型的 q 个参数 $\theta_1,\theta_2,\cdots,\theta_q$ 时，在节 3.3 中我们已经看到，这样将会得到多重解。这些根的每一种线性组合都对应着一个线性表达式，但是，只有一种组合是用过去历史值描述的可逆表示。

例如，考虑 w_t 的 MA 过程

$$w_t=(1-\theta_1 B)a_t$$

假设 $\gamma_0(w)$ 和 $\gamma_1(w)$ 是已知的，我们要推导出 θ_1 和 σ_a^2 。由于

$$\gamma_0=(1+\theta_1^2)\sigma_a^2\ ,\quad \gamma_1=-\theta_1\sigma_a^2\ ,\quad \gamma_k=0\quad (k>1) \tag{6.4.3}$$

于是有

$$-\frac{\gamma_0}{\gamma_1}=\theta_1^{-1}+\theta_1$$

如果 $(\theta_1=\theta,\sigma_a^2=\theta^2)$ 是在给定 γ_0 和 γ_1 时的一组解，那么 $(\theta_1=\theta^{-1},\sigma_a^2=\theta^2\sigma^2)$ 也是解。显然，对于给定的 γ_0 和 γ_1, 存在两个可能的过程：

$$w_t=(1-\theta B)a_t$$

和

$$w_t=(1-\theta^{-1}B)\alpha_t \tag{6.4.4}$$

其中 $\sigma_a^2=\sigma^2\theta^2$ 。若 $-1<\theta<1$, 则 (6.4.4) 就不是可逆表示。尽管如此，该模型还是可以写为

$$w_t=[(1-\theta^{-1}B)(-\theta F)](-\theta^{-1}B\alpha_t)$$

由此，若令 $e_t=-\alpha_{t-1}/\theta$, 则模型成为

$$w_t=(1-\theta F)e_t \tag{6.4.5}$$

其中 e_t 与 a_t 具有相同方差。因此， (6.4.5) 就是简单的“反向”过程，它与下面的正向过程对偶：

$$w_t=(1-\theta B)a_t \tag{6.4.6}$$

正如在 (6.4.6) 中的冲击 a_t 可表示为 w 的当前和过去值的收敛和:

$$a_t = w_t + \theta w_{t-1} + \theta^2 w_{t-2} + \cdots$$

在 (6.4.5) 中的冲击 e_t 表示为 w 的当前和未来值的收敛和:

$$e_t = w_t + \theta w_{t+1} + \theta^2 w_{t+2} + \cdots$$

因此, 根 θ^{-1} 也产生一个可逆过程, 但是, 仅当允许用 w 的未来值表示冲击 e_t 时, 这才是可行. 在表 6.1 中给出的可逆域限定了可接受的参数值, 采用这些参数我们可用既往历史值来表示冲击.

6.4.3 反向过程用于确定初值

假设时间序列 $w_1, w_2, \cdots, w_n$ 获自过程

$$\phi(B)w_t = \theta(B)a_t \tag{6.4.7}$$

在第 7 章中会出现这样的问题: 我们需要去估计 w_0, w_{-1}, w_{-2} 等等, 这一系列值都是在第一个观测值获取之前出现的. 由于在估计模型参数时用到一些基本的递推计算需要用到"初始值", 这时便会遇到上述情况. 假设给定 $w_1, \cdots, w_n$, 我们需要去估计 w_{-l} . 6.4.1 节的讨论表明, 用正向模型 (6.4.7) 或是用反向模型

$$\phi(F)w_t = \theta(F)a_t \tag{6.4.8}$$

都可对 $w_1, \cdots, w_n$ 的概率结构作出同等的解释.

由此可知, w_{-l} 与序列 $w_1, w_2, \cdots, w_n$ 的概率关系和 w_{n+l+1} 与 $w_n, w_{n-1}, w_{n-2}, \cdots, w_1$ 之间的概率关系完全相同. 因此, 为了估计在观测开始之前 $l+1$ 期的值, 我们可以先考虑序列结束后的最优估计或 $l+1$ 期的预报, 然后把这种方法应用到反向序列. 换言之, 就是预测反向序列. 我们称之为"反向预测".

A 6.1 非平稳过程自相关估计值的期望特征

假设 N 个观测值 $z_1, z_2, \cdots, z_N$ 的序列是由下面的非平稳过程产生:

$$\nabla z_t = (1 - \theta B) a_t$$

并计算出其自相关估计值

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{N-k}(z_t - \overline{z})(z_{t+k} - \overline{z})}{\sum_{t=1}^{N}(z_t - \overline{z})^2}$$

通过对该表达式的分子和分母求期望并考虑二者之比, 我们可以得出对自相关估计值特征的一些认识. 依 [199] 我们可以写出

$$\begin{aligned}\mathcal{E}[r_k] &= \frac{E[c_k]}{E[c_0]} \\ &= \frac{\sum_{t=1}^{N-k} E[(x_t - \overline{z})(z_{t+k} - \overline{z})]}{\sum_{t=1}^{N} E[(z_t - \overline{z})^2]}\end{aligned}$$

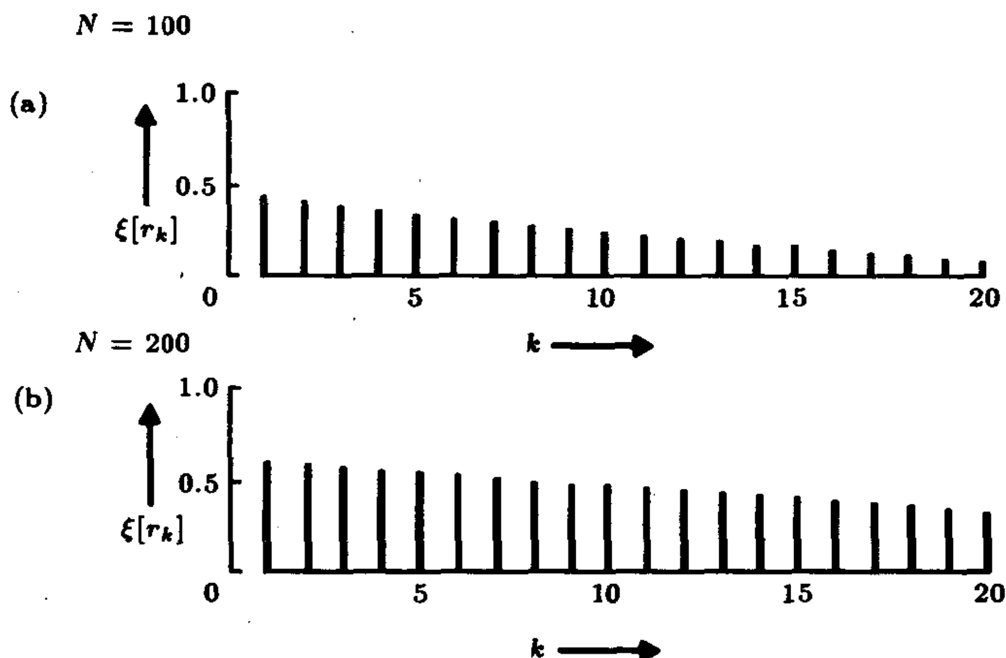


图 A6.1 由 $\nabla z_t=(1-0.8)a_t$ 生成序列的 $\mathcal{E}[r_k]=[c_k]/[c_0]$

经过简单但冗长的代数计算后，我们得到

$$\mathcal{E}[r_k]=\frac{(N-k)[(1-\theta)^2(N^2-1+2k^2-4kN)-6\theta]}{N(N-1)[(N+1)(1-\theta)^2+6\theta]} \quad \text{(A6.1.1)}$$

若 θ 接近于 0, 则 $\mathcal{E}[r_k]$ 将近于 1 . 但是若 θ 值很大，即使对小的 k 值，$\mathcal{E}[r_k]$ 也会比 1 小得多. 在图 A6.1 中给出了 $\theta=0.8$ 时 $N=100$ 和 $N=200$ 的 $\mathcal{E}[r_k]$ 的值，以此来说明上述事实. 正如所料，对于一个非平稳过程而言，虽然期望值之比 $\mathcal{E}[r_k]$ 不会迅速地衰减，但是可以看到即使对较小的滞后它们也不会接近 1.

只要参数值接近于可使模型两边约分，便可得出一个平稳过程，这时体现出类似于平稳化的作用. 例如，前述例子可将模型写为

$$(1-B)z_t=[(1-B)+\delta B]a_t$$

这里，$\delta=0.2$. 当 δ 趋于 0 时，过程的特征将越来越接近于白噪声过程 $z_t=a_t$, 白噪声过程当滞后 $k>0$ 时自相关函数为 0.

A 6.2 得到自回归滑动平均混合模型参数初估计的一般方法

在 6.3 节中我们已经说明了如何得到简单 ARMA 模型参数的初估计. 特别地，在第五部分中的表 B,C 和 D 使得 AR(2),MA(2) 以及 ARMA(1,1) 过程的初估计能很快地被读出. 在本附录中，我们给出获得一般 ARMA(p,q) 过程 $\phi(B)w_t=\theta(B)a_t$ 的初估计的一般方法.

在一般情形下，ARMA(p,q) 过程初估计的计算是基于 $w_t=\nabla^d z_t$ 的前 $p+q+1$ 个自协方差 $c_j\big(j=0,1,\cdots,(p+q)\big)$, 并按以下三个步骤着手.

1. 由自协方差 $c_{q-p+1},\cdots,c_{q+1},c_{q+2},\cdots,c_{q+p}$ 估计自回归参数 $\phi_1,\phi_2,\cdots,\phi_p$.

2. 利用 (1) 中得到的估计值 $\hat{\boldsymbol{\phi}}$ 得到导出序列

$$w_t'=w_t-\hat{\phi}_1w_{t-1}-\cdots-\hat{\phi}_pw_{t-p}$$

计算该序列的前 $q+1$ 个自协方差 $c'_j(j=0,1,\cdots,q)$。

3. 最后，用自协方差 $c'_0,c'_1,\cdots,c'_q$ 来迭代计算滑动平均参数 $\theta_1,\theta_2,\cdots,\theta_q$ 及残差方差 σ_a^2 的估计值.

自回归参数的初估计　利用 (3.4.3) 的结果，自回归参数的初估计的获得可通过下面的 p 个线性方程

$$\begin{aligned}
c_{q+1} &= \hat{\phi}_1 c_q + \hat{\phi}_2 c_{q-1} + \cdots + \hat{\phi}_p c_{q-p+1} \\
c_{q+2} &= \hat{\phi}_1 c_{q+1} + \hat{\phi}_2 c_q + \cdots + \hat{\phi}_p c_{q-p+2} \\
&\vdots \\
c_{q+p} &= \hat{\phi}_1 c_{p+q-1} + \hat{\phi}_2 c_{q+p-2} + \cdots + \hat{\phi}_p c_q
\end{aligned} \tag{A6.2.1}$$

导出滑动平均过程的自协方差　我们现在记 $w'_t=\phi(B)w_t$, 并把该过程处理为下面的滑动平均过程:

$$w'_t=\theta(B)a_t \tag{A6.2.2}$$

首先我们需要用 w_t 的自协方差 c_j 来表示 w'_t 的自协方差 c'_j。可以证明，对 $j=0,1,\cdots,q$ 有

$$c'_j=\sum_{i=0}^{p}\phi_i^2 c_j+\sum_{i=1}^{p}(\phi_0\phi_i+\phi_1\phi_{i+1}+\cdots,+\phi_{p-i}\phi_p)(c_{j+i}+c_{j-i}) \tag{A6.2.3}$$

这里， $\phi_0=-1$。

滑动平均参数的初估计　利用自协方差估计值 c'_j, 通过下面两种迭代过程之一，我们可以得到导出过程 (A6.2.2) 滑动平均参数的初估计。

1. 线性收敛过程。 利用 3.3.2 节给出的 MA 过程自协方差函数的表达式，有

$$\begin{aligned}
\gamma_0 &= (1+\theta_1^2+\cdots+\theta_q^2)\sigma_a^2 \\
\gamma_k &= (-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q)\sigma_a^2 \qquad (k\geq 1)
\end{aligned}$$

我们可以计算出在该确定阶数之下参数 $\sigma_a^2,,\theta_q,\theta_{q-1},\cdots,\theta_1$ 的估计值，只需使用迭代

$$
\begin{aligned}
\sigma_a^2 &= \frac{c_0'}{1+\theta_1^2+\cdots+\theta_q^2} \\
\theta_j &= -\left(\frac{c_j'}{\sigma_a^2}-\theta_1\theta_{j+1}-\theta_2\theta_{j+2}-\cdots-\theta_{q-j}\theta_q\right)
\end{aligned} \tag{A6.2.4}
$$

通常约定 $\theta_0=0$ 。在迭代开始时可令 $\theta_1,\theta_2,\cdots,\theta_q$ 等于零。每次得到 θ_j 和 σ_a^2 的最新值都可以在下一轮计算中使用。例如，在 $q=2$ 的情形，方程 (A6.2.4) 是

$$
\begin{aligned}
\sigma_a^2 &= \frac{c_0'}{1+\theta_1^2+\theta_2^2} \\
\theta_2 &= -\frac{c_2'}{\sigma_a^2} \\
\theta_1 &= -\left(\frac{c_1'}{\sigma_a^2}-\theta_1\theta_2\right)
\end{aligned}
$$

2. 二次收敛过程。 Newton-Raphson 算法是由 Wilson[202] 给出，它比算法 1 的收敛性好。我们记 $\boldsymbol{\tau}'=(\tau_0,\tau_1,\cdots,\tau_q)$，这里

$$
\tau_0^2=\sigma_a^2\,, \qquad \theta_j=\frac{-\tau_j}{\tau_0} \qquad (j=1,2,\cdots,q) \tag{A6.2.5}
$$

若 $\boldsymbol{\tau}'$ 是第 i 次迭到得到 $\boldsymbol{\tau}$ 的估计值，那么，第 $i+1$ 次迭代新的估计值可由下式得出

$$
\boldsymbol{\tau}^{i+1}=(\boldsymbol{\tau})'-(\boldsymbol{T}')^{-1}\boldsymbol{f}_i \tag{A6.2.6}
$$

其中 $\boldsymbol{f}'=(f_0,f_1,\cdots,f_q)$, $f_j=\sum\limits_{i=0}^{q-j}\tau_i\tau_{i+j}-c_j'$, 及

$$
\boldsymbol{T}=\begin{bmatrix}
\tau_0 & \tau_1 & \cdots & \tau_{q-2} & \tau_{q-1} & \tau_q \\
\tau_1 & \tau_2 & \cdots & \tau_{q-1} & \tau_q & 0 \\
\tau_2 & \tau_3 & \cdots & \tau_q & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots \\
\tau_q & 0 & \cdots & 0 & 0 & 0
\end{bmatrix}+\begin{bmatrix}
\tau_0 & \tau_1 & \tau_2 & \cdots & \tau_q \\
0 & \tau_0 & \tau_1 & \cdots & \tau_{q-1} \\
0 & 0 & \tau_0 & \cdots & \tau_{q-2} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & \cdots & \tau_0
\end{bmatrix}
$$

在每一轮迭代中知道了 τ 的值，就可以由 (A6.2.5) 得出参数的值。

例： 考虑下面 ARMA 模型中 ϕ 和 θ 的估计

$$(1-\phi B)\widetilde{z}_t=(1-\theta B)a_t$$

利用 $c_0=1.25, c_1=0.50$ 及 $c_2=0.40$, 实际上，这些值是对应于 $\phi=0.8, \theta=0.5$ 和 $\sigma_a^2=1.00$ 的过程。将 $p=q=1$ 代入 (A6.2.1), 有

$$c_2=\phi c_1$$

可得 ϕ 的估计值，得到 $\hat{\phi}=0.8$ 。因此，利用 (A6.2.3), 导出序列

$$w_t'=\widetilde{z}_t-\phi\widetilde{z}_{t-1}$$

的前两个自协方差是

$$c_0'=(1+\phi^2)c_0-2\phi c_1=1.25$$
$$c_1'=(1+\phi^2)c_1-\phi(c_2+c_0)=-0.50$$

在 (A6.2.4) 中代入这些值，方法 1 的迭代过程是基于下式

$$\sigma_a^2=\frac{1.25}{1+\theta^2}\ ,\qquad \theta=\frac{0.5}{\sigma_a^2}$$

表 A6.1 MA(1) 过程的 σ_a^2 和 θ 初估计值的收敛性

迭代次数	方法 1		方法 2	
	σ_a^2	θ	σ_a^2	θ
0	—	0.000	1.250	0.000
1	1.250	0.400	2.250	0.667
2	1.077	0.464	1.210	0.545
3	1.029	0.486	1.012	0.503
4	1.011	0.494	1.000	0.500
5	1.004	0.498	—	—
6	1.002	0.499	—	—
7	1.001	0.500	—	—
8	1.000	0.500	—	—

类似地，将有关值代入 (A6.2.6)，方法 2 的迭代过程是基于下式

$$\begin{bmatrix} \tau_0^{i+1} \\ \tau_1^{i+1} \end{bmatrix} = \begin{bmatrix} \tau_0^{i} \\ \tau_1^{i} \end{bmatrix} - (\boldsymbol{T}^i)^{-1} \begin{bmatrix} f_0^{i} \\ f_1^{i} \end{bmatrix}$$

其中

$$\begin{bmatrix} f_0^{i} \\ f_1^{i} \end{bmatrix} = \begin{bmatrix} (\tau_0^i)^2 + (\tau_1^i)^2 - c_0' \\ \tau_0^i \tau_1^i - c_1' \end{bmatrix}$$

及

$$\boldsymbol{T}^i = \begin{bmatrix} \tau_0^i & \tau_1^i \\ \tau_1^i & 0 \end{bmatrix} \begin{bmatrix} \tau_0^i & \tau_1^i \\ 0 & \tau_0^i \end{bmatrix} = \begin{bmatrix} 2\tau_0^i & 2\tau_1^i \\ \tau_1^i & \tau_0^i \end{bmatrix}$$

方差 σ_a^2 和 θ 可以由 $\sigma_a^2 = \tau_0^2, \theta = -\tau_1/\tau_0$ 计算出。表 A6.1 列出了方法 1 和 2 迭代收敛的情形。

7 模型的估计

通过识别过程，我们得到了模型的试探形式，随之我们需要得到参数的有效估计。在参数估计出来以后，应对拟合模型进行诊断检验和拟合优度检验。在拟合过程中数据应该被有效地使用，因此，与此相关的拟合优度检验是非常必要的。否则就会发生拟合不当，这完全是由于无效的拟合，而并不是由于模型的形式不恰当。本章包括对随机模型参数估计通常所考虑的极大似然和 Bayes 方法。在本章中，黑体用表示向量和矩阵。因此，$\boldsymbol{X}=\{x_{ij}\}$ 是一个矩阵，其中 x_{ij} 是第 i 行第 j 列的元素，$\boldsymbol{X}'$ 是矩阵 $\boldsymbol{X}$ 的转置。

7.1 似然函数和平方和函数的研究

7.1.1 似然函数

假设有一个 N 个观测值的样本 $\boldsymbol{z}$，我们将它与 N 维随机变量联系起来，已知其概率分布为 $p(\boldsymbol{z}|\boldsymbol{\xi})$，依赖于某些未知参数 $\boldsymbol{\xi}$。我们用向量 $\boldsymbol{\xi}$ 记总的参数集合。具体地，在 ARIMA 模型中它是指 $p+q+1$ 个参数 $(\boldsymbol{\phi},\boldsymbol{\theta},\sigma)$。

在数据取得之前，每一次不同的试验结果对应于一个由参数 $\boldsymbol{\xi}$ 确定的密度 $p(\boldsymbol{z}|\boldsymbol{\xi})$。在数据到手之后，我们观察 $\boldsymbol{\xi}$ 的各种值，希

望能给出实际上得出观测值 z 的那组确定的参数。用于这种目的函数是似然函数 $L(\xi|z)$，它与 $p(z|\xi)$ 具有相同的形式，但在该函数中， z 是固定的，而 ξ 是变化的。我们感兴趣的只是相对值，所以似然函数通常被认为是包含了一项任意的常数乘数。

通常处理对数似然函数 $\ln[L(\xi|z] = l(\xi|z)$ 更为方便，它包含了一项任意的常数加项. 似然函数是估计理论的重要基础，其原因在于由 Fisher[91], Barnard[22] 和 Birnbaum[32] 在某些不同背景下倡导的“似然原理”。该原理指出：(假设模型是正确的) 数据能告诉我们有关参数的所有信息都包含在似然函数中，而数据所有其它方面的东西都是无关紧要的。从 Bayes 的观点来看，似然函数是相当重要的，因为它是参数的后验分布来自于数据的部分。

为了全面了解估计的情况，对似然函数作完整的分析和图形研究是十分必要的，另外，在 Bayes 框架中，我们所考虑情形下参数的后验分布是被似然函数所控制的. 在许多例子中，对于中等或大样本情形，对数似然函数是单峰的，且在靠近最大值的相当大的区域内能用一个二次函数很好地逼近. 这时，对数似然函数可以用其极大值和极大值附近的二阶导数来描述. 使似然函数或对数似然函数达到极大的参数值称为极大似然 (ML) 估计。对数似然函数二阶导数提供了对似然函数“散布程度”的度量，也可用来计算参数的近似标准差。

极大似然估计的极限性质通常是对独立观测而建立的 [164]，但是，正如 Whittle[196] 所述，可将其全盘推广到平稳序列。

在下文的讨论中，我们假定读者熟悉估计理论的某些基本思想，及有关正态分布理论和线性最小二乘的一些重要结果。这些内容是本章需要用到的，我们将在附录 A7.1 和 A7.2 中加以综述。关于时间序列模型参数估计的一些早期重要工作，可参考 [24], [26], [87], [98], [100], [127], [134], [135], [142], [162] ,[163] 和 [197]。

7.1.2 ARIMA 过程的条件似然

假设 $N = n + d$ 个原始观测值 z 构成了一个时间序列，我们记为 $z_{-d+1}, \cdots, z_0, z_1, z_2, \cdots, z_n$。假设该序列是由 (p, d, q) 阶的 ARIMA 过程生成。从这个序列可以产生一个序列 $\boldsymbol{w}$，它是由 $n = N - d$ 个差分值 $w_1, w_2, \cdots, w_n$ 构成，其中 $w_t = \nabla^d z_t$。因此，拟合 ARIMA 模型 (6.1) 参数 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的一般问题等价于拟合序列 $\boldsymbol{w}$ 的平稳可逆[①] ARMA(p, q) 模型的参数，该模型可写为

$$\begin{aligned} a_t = \widetilde{w}_t - \phi_1 \widetilde{w}_{t-1} - \cdots - \phi_p \widetilde{w}_{t-p} \\ + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \end{aligned} \tag{7.1.1}$$

这里，$w_t = \nabla^d z_t, \widetilde{w}_t = w_t - \mu$，且 $E[w_t] = \mu$。

当 $d > 0$ 时，假设 $\mu = 0$ 通常都是正确的 (见 4.1.3 节，6.2.3 节，6.3.5 节及 6.3.8 节的讨论)，否则可用 $\overline{w} = \sum\limits_{t=1}^{n} w_t / n$ 来代替 μ。对于在时间序列分析中通常考虑的样本量而言，这种近似就足够了。但是，如果需要的话，μ 也可以作为一个附加参数加以估计。我们所讨论的方法是对 μ 和其它参数一起作估计。

w 的值不能立刻代入 (7.1.1) 中去计算 a，这是因为差分方程初值所带来的困难。然而，可以认为，w 的 p 个值 $\boldsymbol{w}_*$ 和 a 的 q 个值 $\boldsymbol{a}_*$ 在序列 w 开始之前就给出了。因此，若以这种选择为条件，$a_1, a_2, \cdots, a_n$ 的值可由 (7.1.1) 依次算出。

由此，对于任何给定的参数 $(\boldsymbol{\phi}, \boldsymbol{\theta})$ 选择和初始值 $(\boldsymbol{w}_*, \boldsymbol{a}_*)$，我们可以顺次计算出一组值 $a_t(\boldsymbol{\phi}, \boldsymbol{\theta} | \boldsymbol{w}_*, \boldsymbol{a}_*, \boldsymbol{w}), t = 1, 2, \cdots, n$。现在假设 a 是正态分布，其概率密度为

$$p(a_1, a_2, \cdots, a_n) \propto \sigma_a^{-n} \exp\left[- \left(\sum_{t=1}^{n} \frac{a_t^2}{2\sigma_a^2} \right) \right]$$

对于给定的一组具体数据 $\boldsymbol{w}$，在选择 $(\boldsymbol{w}_*, \boldsymbol{a}_*)$ 的条件下，与参数

[①] 需要特别注意的是，应确保估计值在可逆域内，可参见附录 A7.6。

$(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a)$ 相联系的对数似然函数是

$$l_*(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a^2) = -n\ln(\sigma_a) - \frac{S_*(\boldsymbol{\phi},\boldsymbol{\theta})}{2\sigma_a^2} \tag{7.1.2}$$

在上式中，按照前面的讨论，不必包括附加的常数项，且有

$$S_*(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=1}^{n} a_t^2(\boldsymbol{\phi},\boldsymbol{\theta}|\boldsymbol{w}_*,\boldsymbol{a}_*,\boldsymbol{w}) \tag{7.1.3}$$

在上面各式中，似然函数和平方和函数所用的星号下标是为了强调这些值是以初值的选择为条件的。我们注意到，条件对数似然函数 l_* 仅通过条件平方和函数而与数据相联系。由此可得，在空间 $(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a)$ 中对于 σ_a^2 的任一确定值，l_* 的等值线就是 S_* 的等值线，于是极大似然估计与最小二乘估计相同，且在正态假定下，我们一般可通过研究条件平方和函数的性质来研究条件似然函数的性质。特别当 σ_a^2 任意固定时，l_* 是 S_* 的线性函数。通过使条件平方和函数 $S_*(\boldsymbol{\phi},\boldsymbol{\theta})$ 极小化而得到的参数估计就称为条件最小二乘估计

7.1.3 对于条件计算的初值选择

我们将简要地讨论非条件似然函数的计算，严格说来，这是我们进行参数估计所需要的。然而在某些应用中，当 n 为适中或较大时，在 (7.1.3) 中用适当的值代替 $\boldsymbol{w}_*$ 和 $\boldsymbol{a}_*$ 的元素，利用该条件似然函数可以得到非条件似然函数的充分近似。一种方法就是设 $\boldsymbol{w}_*$ 和 $\boldsymbol{a}_*$ 的元素等于它们的非条件期望。$\boldsymbol{a}_*$ 中元素的非条件期望为 0, 且如果模型不含确定性部分，并且特别当 $\mu=0$ 时，$\boldsymbol{w}_*$ 中元素的非条件期望也为 0[①] 但是，如果 $\phi(B)=0$ 的一些根接近单位圆的边界，过程近似为非平稳，这种近似就可能很不好。在这种情形下，起始的数据值 w_1 可能偏离非条件期望相当远，

① 如果假设 $Ew_t=\mu\neq0$ 是合理的，则可用 $\overline{w}$ 来代替 $\boldsymbol{w}_*$ 的每个元素。

这种初值的引入可能会产生巨大的瞬变状态，而瞬变的消失又是缓慢的. 在拟合 (p,d,q) 阶模型时，我们有时将使用的一种更可靠的方法是：利用 (7.1.1) 从 a_{p+1} 开始向前计算 a, 且令既往的 a 都为零. 因此对于 w 来说，用到的都是实际发生的值.

利用这种方法，我们可以对 a_t 的 $n-p=N-p-d$ 个值的平方求和，当然，对于较长的序列来说，这种信息的轻微损失是不足道的。在没有自回归项的情形这两种方法是等价的. 对于第 9 章中讨论的季节序列来说，条件近似是不能令人满意的，而条件计算就变得更加必要了。下面我们将通过简单的例子来说明条件平方和 S_* 的递推计算。

对 (0,1,1) 过程条件平方和的计算　序列 B 在表 6.4 中已被初步识别为是 IMA(0,1,1) 过程：

$$\nabla z_t=(1-\theta B)a_t \qquad (-1<\theta<1) \tag{7.1.4}$$

即

$$a_t=w_t+\theta a_{t-1}$$

这里，$w_t=\nabla z_t$, 及 $E[w_t]=0$. 我们来回忆一下，在第 6 章中我们曾得到初步的矩估计 (表 6.5) 并由数据给出 θ 的值接近于零。对于特定的参数值 $\theta=0.5$, 在表 7.1 中列出了前几个 a 的计算结果. a 由 $a_t=w_t+0.5a_{t-1}$ 递推计算，且精确到两位小数。与上面的讨论一致，为了着手计算，a_0 被设为 0. 这个值用黑体表示. 如此继续，我们求出

$$S_*(0.5)=\sum_{t=1}^{368}a_t^2(\theta=0.5|a_0=0)=27,694$$

用从 $\theta=-0.5$ 到 $\theta=+0.5$ 步长为 0.1 的 θ 值，在表 7.2 中列出了条件平方和 $S_*(\theta)$(给定 $a_0=0$ 时) 的计算值。

7.1.4 非条件似然；平方和函数；最小二乘估计

在附录 A7.3 中证明了：假设 $N = n + d$ 个观测是由一个 ARIMA 过程产生的，那么，对应的非条件对数似然函数由下式给出

$$l(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a) = f(\boldsymbol{\phi},\boldsymbol{\theta}) - n\ln(\sigma_a) - \frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{2\sigma_a^2} \tag{7.1.5}$$

其中 $f(\boldsymbol{\phi},\boldsymbol{\theta})$ 是 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的函数。非条件平方和函数为

$$S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=1}^{n}[a_t|\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}]^2 + [\boldsymbol{e}_*]'\boldsymbol{\Omega}^{-1}[\boldsymbol{e}_*] \tag{7.1.6}$$

这里 $[a_t|\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}]$ 表示 a_t 以 $\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}$ 为条件的条件期望。当不至混淆时，我们将把条件期望简记为 $[a_t]$ 。在 (7.1.6) 中我们用 $\boldsymbol{e}_* = (w_{1-p},\cdots,w_0,a_{1-q},\cdots,a_0)'$ 表示 $p+q$ 个初始值的向量，这是在时刻 $t=1$ 之前过程 w_t 和 a_t 所需要的， $\boldsymbol{\Omega}\sigma_a^2 = \text{cov}(\boldsymbol{e}_*)$ 是 $\boldsymbol{e}_*$ 的协方差阵， $[\boldsymbol{e}_*] = ([w_{1-p}],\cdots,[w_0],[a_{1-q}],\cdots,[a_0])'$ 表示在给定 $\boldsymbol{w},\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 时初始值的条件期望 (“反向预报 ”) 向量。另一种表示平方和的方法是 $S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=-\infty}^{n}[a_t]^2$, 与 (7.1.6) 相比，有 $\sum_{t=-\infty}^{0}[a_t]^2 = [\boldsymbol{e}_*]]\boldsymbol{\Omega}^{-1}[\boldsymbol{e}_*]$.

表 7.1 取 θ=0.5, 对序列 B 前 10 个 a 值的递推计算

l	z_t	$w_t=\nabla z_t$	$a_t=w_t+0.5a_{t-1}$
0	460		0
1	457	−3	−3.00
2	452	−5	−6.50
3	459	7	3.75
4	462	3	4.88
5	459	−3	−0.56
6	463	4	3.72
7	479	16	17.86
8	493	14	22.93
9	409	−3	8.46

表 7.2 对序列 B 拟合 $\nabla z_t=(1-\theta B)a_t$ 模型的平方和函数

θ	$\lambda=(1-\theta)$	$S_*(\theta)$	$S(\theta)$	θ	$\lambda=(1-\theta)$	$S_*(\theta)$	$S(\theta)$
–0.5	1.5	23,929	23,928	0.1	0.9	19,896	19,896
–0.4	1.4	21,595	21,595	0.2	0.8	20,851	20,851
–0.3	1.3	20,222	20,222	0.3	0.7	22,315	22,314
–0.2	1.2	19,483	19,483	0.4	0.6	24,471	24,486
–0.1	1.1	19,220	19,422	0.5	0.5	27,694	27,691
0.0	1.0	19,363	19,363				

通常，$f(\boldsymbol{\phi},\boldsymbol{\theta})$ 只对较小的 n 有意义，对于中等或较大的 n 值，(7.1.5) 被 $S(\boldsymbol{\phi},\boldsymbol{\theta})/2\sigma_a^2$ 所控制，因此在参数空间 $(\boldsymbol{\phi},\boldsymbol{\theta})$ 中，非条件平方和函数的等值线极其接近似然和对数似然函数的等值线。特别地，通过极小化平方和函数 (7.1.6) 所得到的参数估计，我们称之为 (非条件或严格的) 最小二乘估计，它通常可以作为极大似然估计很接近的近似值。从 Bayes 观点来看，在 7.4 节讨论的假设之下，对于所有的 AR(p) 和 MA(q) 后验密度基本上只是 $S(\boldsymbol{\phi},\boldsymbol{\theta})$ 的函数。因此，最小二乘估计非常接近具有极大后验密度的估计。在本节的剩余部分和 7.1.5 节，我们着重强调 (7.1.6) 所定义的非条件平方和函数 $S(\boldsymbol{\phi},\boldsymbol{\theta})$ 的计算、研究和应用，以及有关最小二乘估计的计算。

在非条件平方和的计算中，$[a]$ 的值是通过在 (7.1.6) 中取条件期望，然后再进行递推计算。预先的反向计算给出了 $[w_{-j}]$ 和 $[a_{-j}], j=0,1,2,\cdots$ 的值 (即反向预报值)，这些值是着手向前递推所需要的。

滑动平均过程非条件平方和的计算　为了说明起见，我们考虑 IBM 股票价格的例子，并再次使用在表 7.1 中给出的该序列的前 10 个值[1]。用 $w_t=\nabla z_t$, 我们在 6.2 节中已看到，识别的模型是 (0,1,1) 阶。因此，由 (7.1.6) 可知，$S(\theta)$ 需要的唯一反向预报值是 $[a_0]$。我们通过描述一个近似但仍是准确的方法入手，以期

[1] 当然，在实际中，有用的参数估计不可能从这样少的观测值得到，我们利用这组数据只是为了说明计算。

得到 $[a_0]$ 的值，该方法利用了 w_t 向后模型的有趣特征且方法本身就是具有启发性的。我们回忆一下 6.4.3 节，关于 w_t 的模型可以写为向前或向后形式

$$w_t = (1-\theta B)a_t \ , \qquad w_t = (1-\theta F)e_t$$

这里，再次假设 $\mu = E[w_t]$ 等于零。因此，我们可写为

$$[e_t] = [w_t] + \theta[e_{t+1}] \tag{7.1.7}$$

$$[a_t] = [w_t] + \theta[a_{t-1}] \tag{7.1.8}$$

这里，$[w_t] = w_t, t = 1, 2, \cdots, n$, 且对 $t < 0$ 时是 w_t 的反向预报值。这是我们在计算中所需要的两个基本方程。表 7.3 给出了方便的计算格式，我们从在表中列入已知的数据开始：

1. 数据 $z_0, z_1, \cdots, z_{10}$, 由此我们可计算出一阶差分 $w_1, w_2, \cdots, w_9$;

2. $[e_0], [e_{-1}], \cdots$ 的值都是 0, 这是因为 $e_0, e_{-1}, \cdots$ 依分布独立于 $\boldsymbol{w}$。

3. $[a_{-1}], [a_{-2}], \cdots$ 的值都是 0, 因为对于任何 MA(q) 过程 $a_{-q}, a_{-q-1}, \cdots$ 依分布独立于 $\boldsymbol{w}$。但是应当注意，通常 $[a_0], [a_{-1}], \cdots, [a_{-q+1}]$ 是非零的，需由反向预报得出。所以，在本例中 $[a_0]$ 就是这样得到的。

从序列的尾部开始，(7.1.7) 现在被用来计算 $[e_t], t = 9, 8, 7, \cdots, 1$。如前文所述计算条件平方和的方法，我们也作同样的近似，开始着手反向过程。在目前情形下，这相当于令 $[e_{10}] = 0$。通常，这种近似的效果是将一种瞬时状态引入系统，由于 $\phi(B)$ 和 $\theta(B)$ 是稳态算子，对于中等长度的序列来说，在开始时刻达到的瞬态几乎肯定是可以忽略的，因此不会影响 a 的计算。正如后文所述，如果需要的话，这种近似的合理性在任合情形下都可以通过两轮迭代来检验。

在表 7.3 中，为了开始着手递推，对应于 $t = 9$ 的那一行上，在第 6 列对未知值 $0.5[e_{10}]$ 我们写上一个 0(用黑体表示)。因此，

表 7.3 由序列 B 的前 10 个值计算 $[a], \theta=0.5$

t	z_t	$[a_t]$	$0.5[a_{t-1}]$	$[w_t]$	$0.5[e_{t+1}]$	$[e_t]$	第二轮 $[e_t]$	u_t
−1	[458.4]	0	0	0	0	0	0	
0	460	1.6	0	1.6	−1.6	0	0	−2.1
1	457	−2.2	0.8	−3.0	−0.1	−3.1	−3.1	−4.1
2	452	−6.1	−1.1	−5.0	4.8	−0.2	−0.2	−2.3
3	459	3.9	−3.0	7.0	2.6	9.6	9.6	8.5
4	462	5.0	2.0	3.0	2.3	5.3	5.2	9.5
5	459	−0.5	2.5	−3.0	7.6	4.6	4.4	9.2
6	463	3.7	−0.2	4.0	11.1	15.1	14.9	19.4
7	479	17.9	1.9	16.0	6.2	22.2	21.7	31.4
8	493	22.9	9.0	14.0	−1.5	12.5	11.4	27.5
9	490	8.5	11.5	−3.0	0	−3.0	−5.1	8.5
							−4.2	

利用 (7.1.7) 我们得到

$$[e_9] = [w_9] + 0.5[e_{10}] = w_9 + 0 = -3$$

于是 $0.5[e_9] = -1.5$ 可以写到 $t = 8$ 的一行上，这样，我们便能去计算 $[e_8]$, 依此类推。最后，我们得到

$$[e_0] = [w_0] + \theta[e_1]$$

即

$$0 = [w_0] - 1.6$$

由此得出 $[w_0] = 1.6$, 从而有 $[w_{-h}] = 0, h = 1, 2, 3, \cdots$ 。

利用 (7.1.8) 及 $t = 0$, 我们得到

$$[a_0] = [w_0] + \theta[a_{-1}] = 1.6 + (0.5)(0) = 1.6$$

然后我们又可继续向前计算剩下的 $[a]$ 值，由此得出 $S(0.5) = \sum_{t=0}^{9}[a_t|0.5, \boldsymbol{w}]^2 = 1016.406$ 。对比表 7.1 和 7.3 给出的 $[a]$ 值，表明在这个具体例之中，由于初值的改变而引起的瞬变当 $t > 5$ 时几乎没有什么影响。

与上述近似方法相比较，附录 7.3 事实上给出了对于一般情形计算 $[a]$, 从而计算 $S(\theta)$ 的精确方法。对于 w_t 的 AR(1) 模型来说，该方法首先涉及计算 $a_t(a_0=0)$ 的值，我们简记为 a_t^0, 由 7.1.3 节讨论的条件方法，有

$$a_t^0 = w_t + \theta a_{t-1}^0 \qquad (t=1,2,\cdots,n)$$

取 $a_0^0=0$ 作为初值。然后采用反向递推来求得 $u_t = a_t^0 + \theta u_{t+1}$, 从 $t=n$ 开始，直到 $t=0$, 以 $u_{n+1}=0$ 作为初值。最后，精确的反向预报 $[a_0]$ 由 $[a_0] = -u_0(1-\theta^2)/(1-\theta^{2/(n+1)})$ 给出，利用这一初值，如 7.1.3 那样， $[a_t]$ 可以用下面的正向递推算出：

$$[a_t] = w_t + \theta[a_{t-1}] \qquad (t=1,2,\cdots,n)$$

精确的平方和为 $S(\theta) = \sum_{t=0}^{n}[a_t]^2$ 。为了说明和对比，利用表 7.1 中的条件值 a_t^0, 我们由反向递推得出 $t=9,8,\cdots,0$ 的 u_t 值，并列在表 7.3 的最后一列。因此我们求得 a_0 精确的反向预报值为 $[a_0] = -u_0(1-\theta^2)/(1-\theta^{2(n+1)}) = 1.549$, 这与由反向模型方法得出的近似值 1.545 非常相近，且这种微小的差别对其余 $[a_t]$ 的计算基本上没有影响。

在实际中，可以使用近似反向模型计算的两轮迭代方法，但这几乎总是不必要的。然而，我们只是通过 MA(1) 模型的 9 个观测值构成的非真实的短序列来说明这一方法。为此，我们采用在第一轮迭代中计算的 $[a_9]=8.47$ 为 a_9 的值，并以此作为新一轮迭代的开始，用 (7.1.8) 计算 w_{10} 的正向预报 $[w_{10}] = -\theta[a_9] = -4.23$ 。然后可将该值代入反向方程，得到第一个 $[e_{10}] = [w_{10}] = -4.23$(因为对于反向模型 $[e_{11}]=0$), 最后，求得新的反向预报值 $[w_0]$, 从而得到新的 $[a_0]=[w_0]$, 由此而得出 $[a_t]$ 的新值。反向递推 $[e_t]$ 的第二轮结果在表 7.3 的倒数第二列给出。第二论计算最后得出 $[a_0]=1.549$, 该值非常接近于反向模型第一轮计算所得的值，且基本上等于精确值。因此， $[a]$ 的进一步计算将会得到如前一样的相同结果。

进一步，由过程的可逆性质，通常 $S(\boldsymbol{\phi},\boldsymbol{\theta}|\boldsymbol{w})$ 可由 $[e_t]$ 平方和的计算给出。利用这一事实，对 MA(1) 示例的第二轮结果，我们由表 7.3 求出

$$S(0.5)=\sum_{t=1}^{10}[e_t]^2=1016.406$$

这与由 $[a_t]$ 求得的值一致。

对整个序列用前面的方式处理，并采用精确的方法，我们求出非条件平方和为

$$S(0.5)=\sum_{t=0}^{368}[a_t|0.5,\boldsymbol{w}]^2=27,691$$

对于这个具体的例子，该值非常接近于相应的条件值 $S_*(0.5)=27,694$ 。表 7.2 给出了 θ 在 -0.5 到 +0.5 之间的非条件平方和 $S(\theta)$, 对于这个具体的例子，这些值非常接近于条件值 $S_*(\theta)$ 。

7.1.5 计算非条件平方和的一般方法

在上面的例子中，w_t 是一个均值为零的一阶滑动平均过程。由此得出，提前期大于 1 的所有预报值都是 0, 因此，为了用近似的反向模型方法着手进行递推计算只需要一个初始值 (反向预报值 $[w_0]=1.6$), 在精确方法中也只需要一个反向预报值 $[a_0]$ 。对于一个 q 阶滑动平均过程则需要 q 个非零初始值 $[w_0],[w_{-1}],\cdots,[w_{1-q}]$, 或者等价地，在精确方法中对 $S(\boldsymbol{\theta})=\sum\limits_{t=1-q}^{n}[a_t]^2$ 需要 q 个值 $[a_0],[a_{-1}],\cdots,[w_{1-q}]$ 。我们在 7.3.1 节中讨论的特殊方法对于估计自回归模型的参数也是可行的。然而我们在附录中证明了，本节所述方法对于任意 ARIMA 模型都能给出非条件平方和。

特别地，假设 w_t 是由一个平稳的正向模型产生

$$\phi(B)\widetilde{w}_t=\theta(B)a_t \tag{7.1.9}$$

其中 $\nabla^d z_t = w_t, \widetilde{w}_t = w_t - \mu$. 那么，它们也可以由反向模型所产生

$$\phi(F)\widetilde{w}_t = \theta(F)e_t \tag{7.1.10}$$

同前面一样，在运用反向模型的近似方法中，我们可以首先利用 (7.1.10) 给出反向预报值 $[\widetilde{w}_{-j}|\boldsymbol{w}, \boldsymbol{\phi}, \boldsymbol{\theta}]$ 。从理论上讲，自回归算子的存在使得这种估计值的序列是无穷多项。然而，由于该算子的平稳性，实际中对于中等大小的 Q, 在点 $t = -Q$ 以外，估计值 $[\widetilde{w}_t]$ 基本上等于零。

因此，对充分的近似值，我们可以写为

$$\widetilde{w}_t = \phi^{-1}(B)\theta(B)a_t = \sum_{j=0}^{\infty}\psi_j a_{t-j} \simeq \sum_{j=0}^{Q}\psi_j a_{t-j}$$

这意味着原来的混合过程可以用一个 Q 阶滑动平均过程来代替，因此，在 7.1.4 节中针对滑动平均过程所归纳的方法就可以使用。

于是，通常在 (7.1.10) 和 (7.1.9) 中取条件期望，可以得到产生条件期望 $[a_t|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{w}]$ 的对偶方程组，即

$$\phi(F)[\widetilde{w}_t] = \theta(F)[e_t] \tag{7.1.11}$$

用来生成反向预报值，而

$$\phi(B)[\widetilde{w}_t] = \theta(B)[a_t] \tag{7.1.12}$$

用来产生 $[a_t]$ 。如果我们发现在某个领先期之外预报值的量级是可以忽略的，则递推计算在向前进行时有

$$\begin{aligned} [e_{-j}|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{w}] &= 0 \qquad (j = 0, 1, 2\cdots) \\ [a_{-j}|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{w}] &= 0 \qquad (j > Q - 1) \end{aligned} \tag{7.1.13}$$

且平方和近似为 $S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum\limits_{t=1-Q}^{n}[a_t]^2$. 如前所述，如果需要的话可以采用该近似方法中的第二轮迭代。

然而，如附录 A7.3 中所述，对一般的模型 (7.1.9) 可以使用精确方法求得平方和为

$$S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=1-q}^{n} [a_t]^2 + ([\boldsymbol{w}_*] - \boldsymbol{C}'[\boldsymbol{a}_*])'\boldsymbol{K}^{-1}([\boldsymbol{w}_*] - \boldsymbol{C}'[\boldsymbol{a}_*]) \quad (7.1.14)$$

这里，向量 $[\boldsymbol{w}_*]' = ([w_{1-p}], \cdots, [w_0])$ 和 $[\boldsymbol{a}_*]' = ([a_{1-q}], \cdots, [a_0])$ 是由 (A7.3.12) 求得的精确反向预报值。它们由 $[\boldsymbol{e}_*] = ([\boldsymbol{w}_*]', [\boldsymbol{a}_*]')] = \boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{u}$ 给出，其中向量 $\boldsymbol{u}$ 的值 $u_t, t = 1, \cdots, n$ 通过反向递推 $u_t = a_t^0 + \theta_1 u_{t+1} + \cdots, +\theta_q u_{t+q}$ 并取初值 $u_{n+1} = \cdots = u_{n+q} = 0$ 而得到，这里的 a_t^0 是由 (7.1.12) 计算的 a_t 的条件值，且取零初值 $a_{1-q}^0 = \cdots = a_0^0 = 0, w_{1-p}^0 = \cdots = w_0^0 = 0$ 。于是精确的 $[a_t]$ 可由下面类似的递推方法算出

$$\begin{aligned}[a_t] = [w_t] - \phi_1[w_{t-1}] - \cdots - \phi_p[w_{t-p}] + \theta_1[a_{t-1}] \\ + \cdots + \theta_q[a_{t-q}]\end{aligned} \quad (7.1.15)$$

对 $t = 1, 2, \cdots, n$, 以精确的反向预报值作为初值，这里，$[w_t] = w_t, 1 < t < n$ 。(7.1.14) 的计算中所必须的矩阵 $\boldsymbol{C}, \boldsymbol{K}, \boldsymbol{D}$ 和 $\boldsymbol{F}$ 都在附录 A7.3 中清楚地给出了定义。

自回归滑动平均混合过程非条件和的计算　为了说明起见，考虑下面 $n = 12$ 个 $w_t = \nabla z_t$ 的顺序值：

t	1	2	3	4	5	6	7	8	9	10	11	12
w_t	1.1	4.3	3.0	–0.7	1.6	3.2	0.3	–1.9	–0.3	–0.3	0.8	2.0

假设我们希望计算与 ARIMA$(1, d, 1)$ 过程相联系的非条件平方和，该过程为

$$(1 - \phi B)w_t = (1 - \theta B)a_t$$

$\nabla^d z_t = w_t$, 且假定 w 有零均值。当然，建立在 12 个观测值上的估计几乎是没有价值的，但是，这个短序列仅仅是用于解释计算

的性质。我们用参数 $\phi=0.3,\theta=0.7$ 作为例子加以说明。于是 (7.1.15) 可以记为

$$[a_t]=[w_t]-0.3[w_{t-1}]+0.7[a_{t-1}] \tag{7.1.16}$$

其中 $[w_t]=w_t(t=1,2,\cdots,n)$。

在表 7.4 中给出了必要的计算过程。先把数据写在中央的一列上，然后把条件初值 w_0^0,a_0^0 写上。现在从正向方程 (7.1.16) 着手计算，用 7.1.3 节中所述有关条件计算的精确方式得到 a_t^0 为 $a_t^0=w_t-\phi w_{t-1}^0+\theta a_{t-1}^0$，其中 $w_t^0=w_t, 1<t<n$。这样，我们在 $t=1$ 的一行中用 0 来代替 $-0.3w_{t-1}^0$ 和 $0.7a_{t-1}^0$，得到 $a_1^0=w_t$。现在先运用 (7.1.16) 作递推计算求得 a_t^0。紧接着从 $u_{13}=0$ 开始通过反向递推得到值 $u_t=a_t^0+0.7u_{t+1}$，因此在 $t=12$ 的一行上先把 $0.7u_{t+1}$ 的值写为 0. 然后通过解方程 $\boldsymbol{D}[\boldsymbol{e}_*]=\boldsymbol{F}'\boldsymbol{u}=[\phi,-\theta]'u_1$ 可以得到精确的反向预报值 $[\boldsymbol{e}_*]'=([w_0],[a_0])$，

表 7.4 $[a]$ 的平方和 $S(0.3,0.7)$ 的计算，数据是由过程 $(1\text{-}0.3B)w_t=(1\text{-}0.7B)a_t$ 所生成序列中的 12 个值

t	a_t^0	$0.7a_{t-1}^0$	$-0.3w_{t-1}^0$	$[w_t]$	u_t	$0.7u_{t+1}$	$0.7[a_{t-1}]$	$[a_t]$
0	0			−2.46				−2.84
1	1.10	0	0	1.1	9.53	8.43	−1.99	−0.15
2	4.74	0.77	−0.33	4.3	12.05	7.31	−0.10	3.87
3	5.03	3.32	−1.29	3.0	10.44	5.41	2.71	4.42
4	1.92	3.52	−0.90	−0.7	7.73	5.81	3.09	1.49
5	3.15	1.34	0.21	1.6	8.30	5.15	1.04	2.85
6	4.93	2.21	−0.48	3.2	7.36	2.43	2.00	4.72
7	2.79	3.45	−0.96	0.3	3.47	0.68	3.30	2.64
8	−0.04	1.59	−0.09	−1.9	0.98	1.01	1.85	−0.14
9	0.24	−0.03	0.57	−0.3	1.45	1.21	−0.10	0.17
10	−0.04	0.17	0.09	−0.3	1.72	1.76	0.12	−0.09
11	0.86	−0.03	0.09	0.8	2.52	1.65	−0.06	0.83
12	2.36	0.60	−0.24	2.0	2.36	0	0.58	2.34

在这个例子中它们是

$$\begin{bmatrix} 5.864 & -6.009 \\ -6.009 & 7.684 \end{bmatrix} \begin{bmatrix} [w_0] \\ [a_0] \end{bmatrix} = \begin{bmatrix} 0.3 \\ -0.7 \end{bmatrix} u_1$$

得到 $[w_0] = -0.258u_1 = -2.462, [a_0] = -0.297u_t = -2.836$ 。现在开始最后的正向递推，利用 (7.1.16) 并用反向预报值作为初值，可计算出 $[a_t]$ 。在表 7.4 中，计算结果给出了两位小数。

用附录 (A7.3) 中的 (A7.3.14)(对于 ARMA(1,1) 的特殊情形也可用 (A7.3.19)) 及前文所述内容可以得出非条件平方和 $S(\phi, \theta)$ 。因此，由于在 (7.1.14) 中 $\boldsymbol{C} = E[a_0 w_0]/\sigma_a^2 = 1$, 便有

$$S(0.3, 0.7) = \sum_{t=0}^{12} [a_t]^2 + \frac{([w_0] - [a_0])^2}{\boldsymbol{K}} = 89.158$$

这里 $\boldsymbol{K} = \sigma_a^{-2}\gamma_0 - 1 = (\phi - \theta)^2/(1 - \phi^2)$ 。

在对 IBM 的序列 B 拟合 (0,1,1) 过程时我们看到，条件平方和给出了对于非条件平方和非常接近的近似。在当前这个例子中可以看到，情况并不总是这样。

我们在 7.1.3 节曾提到，有两种平方和可以作为非条件平方和的近似。可用下面方法求得：(1) 从第一个有效的观测值开始着手递推，令所有未知的 a 为 0, 所有的 w 等于其非条件期望；(2) 从第 p 个观测值开始着手递推，只使用观测到的 w 值，并令未知的 a 为 0. 这是两种条件平方和，将它们与由 $[a_t]$ 求得的非条件平方和 (3) 作如下比较。

1. 未知的 w 和 a 令其为 0, 有：

$$\sum_{t=1}^{12} (a_t | 0.3, 0.7, w_0 = 0, a_0 = 0, \boldsymbol{w})^2 = 101.044$$

2. 未知的 a 设为 0, 有：

$$\sum_{t=2}^{12} (a_t | 0.3, 0.7, w_1 = 1.1, a_1 = 0, \boldsymbol{w})^2 = 82.441$$

3. 非条件计算结果是:

$$\sum_{t=1}^{12}[a_t|0.3,0.7,\boldsymbol{w}]^2+\frac{([w_0]-[a_0])^2}{K}=89.158$$

(1) 中的平方和对 (3) 的近似很差, 在这 12 个数据的序列中误差超过了 10%, 当然, 如果序列长一些, 误差会有所降低. 这种情形是由于初值选择所导致的瞬变状态在经过 12 个值后会基本消失, 在表 7.4 的结果中我们能看到这一点。另一方面, 如果 (2) 中的条件平方和所允许的平方个数是 11 个而不是 12 个, 那么, 看来 (2) 能给出比 (1) 更令人满意的近似。基于上述理由, 如果要使用条件近似, 那么通常用形式 (2) 而不是 (1) 。然而, 正如第 9 章将要讨论的, 对于季节序列而言, 条件近似不能令人满意, 通常应计算非条件平方和。

当然, 我们最终的关注焦点仍然是使平方和达到极小 (或似然函数达极大) 的参数值, 即使近似的平方和与精确的平方和函数有所不同, 但最小平方 (或最大似然) 估计值可能非常接近。Dent 和 Min[80] 以及 Ansley 和 Newbold[14] 进行了模拟研究, 从经验上考察和比较了 ARMA 模型的条件最小平方、非条件最小平方以及极大似然估计。通常对于大样本情形, 条件与非条件的最小平方估计可以作为极大似然估计的满意近似。然而, 模拟结果表明, 对于小样本或中等样本情形应选择极大似然估计, 特别对于滑动平均算子有一个根接近可逆域边界的情形 Hillmer 和 Tiao[110], Osborn[152] 给出不同估计值特性的一些补充信息, 他们把 MA(1) 模型的条件平方和、非条件平方和及似然函数的期望值作为未知参数 θ 的函数, 对于不同样本容量 n 进行了考察。这些研究工作提出一些思想, 揭示了对于不同的样本容量相应的估计量会表现出怎样的特征, 这些结果与模拟研究的结果是一致的。

7.1.6 最小平方和函数的图形研究

表 7.2 所给 IBM 数据平方和函数 $S(\theta)$ 的图形在图 7.1 中绘出。全局最小平方和约为 $\theta=-0.09$ ($\lambda=1.09$)，它是最小二乘估计 (在正态假定之下)，且是参数 θ 极大似然估计的极好近似。

平方和函数的图形研究很容易推广到两参数情形，只需在适当的参数网格点上计算平方和函数，并画出等值线. 正如我们在 7.1.4 节所讨论的，在正态性假设下，这种等值线非常接近于似然函数的等值线. 对于多数实际问题的要求，在计算机输出的格点上凭眼睛描出粗略的等值线就足够了。然而，由计算机的输出自动作图的设备也是常见的. 用这些方法，由格点值画出等值线图是很容易处理的。图 7.2 给出了序列 B 平方和函数 $S(\lambda_0,\lambda_1)$ 的格点值，拟合的 IMA(0,2,2) 过程为

$$\begin{aligned}\nabla^2 z_t &= (1-\theta_1 B-\theta_2 B^2)a_t \\ &= [1-(2-\lambda_0-\lambda_1)B-(\lambda_0-1)B^2]a_t\end{aligned} \tag{7.1.17}$$

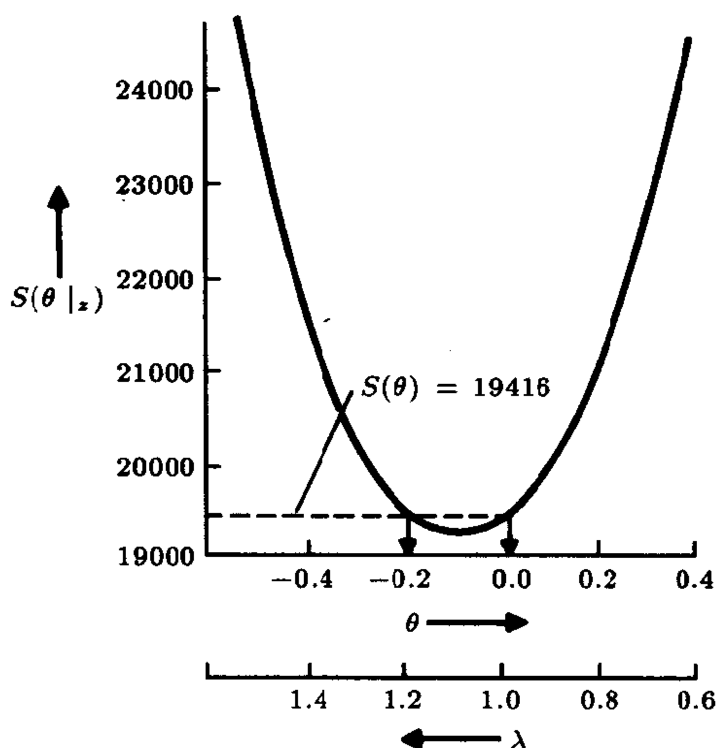


图 7.1 序列 B 的 $S(\theta)$ 图形

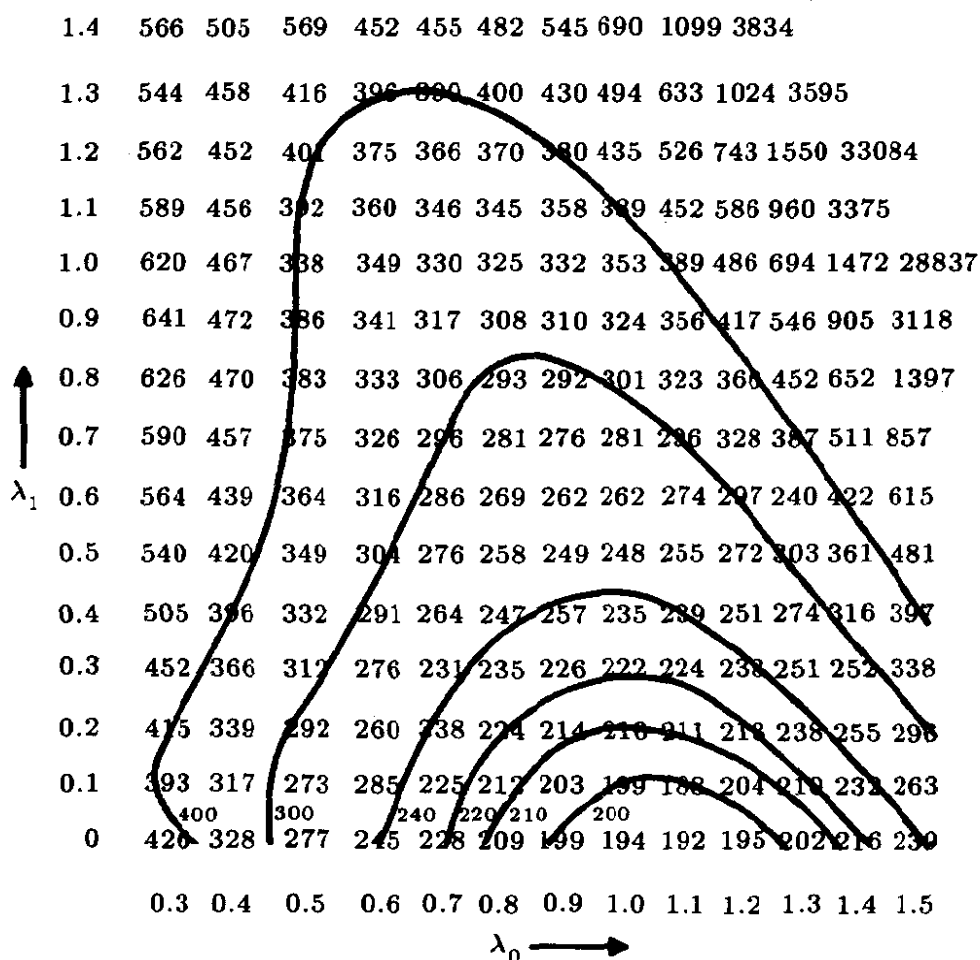


图 7.2 序列 B 在 (λ_0,λ_1) 的格点上的 $S(\lambda_0,\lambda_1)\times10^{-2}$ 值和近似的等值线

或表示为

$$\nabla^2 z_t = (\lambda_0\nabla + \lambda_1)a_{t-1} + \nabla^2 a_t$$

在图 7.2 中最小平方和约在 $\hat{\lambda}_0 = 1.09, \hat{\lambda}_1 = 0.0$ 处。因此由图中确认在此例中应选的模型应是 IMA(0,1,1) 过程. 这里所说明的拟合模型的方式比预料更精细，我们可以由此提供关于最初识别的有用的确认。当然，应该在我们所“担忧 ”的那些必要的方向上进行模型的进一步精细化。

三个参数　当我们希望研究三个参数联合估计的情形，我们可以对第三个参数的多个值画出二维的等值线图。为了说明问题，在图 7.3 中给出了序列 A,C 和 D 一系列这样图形中的一部分。

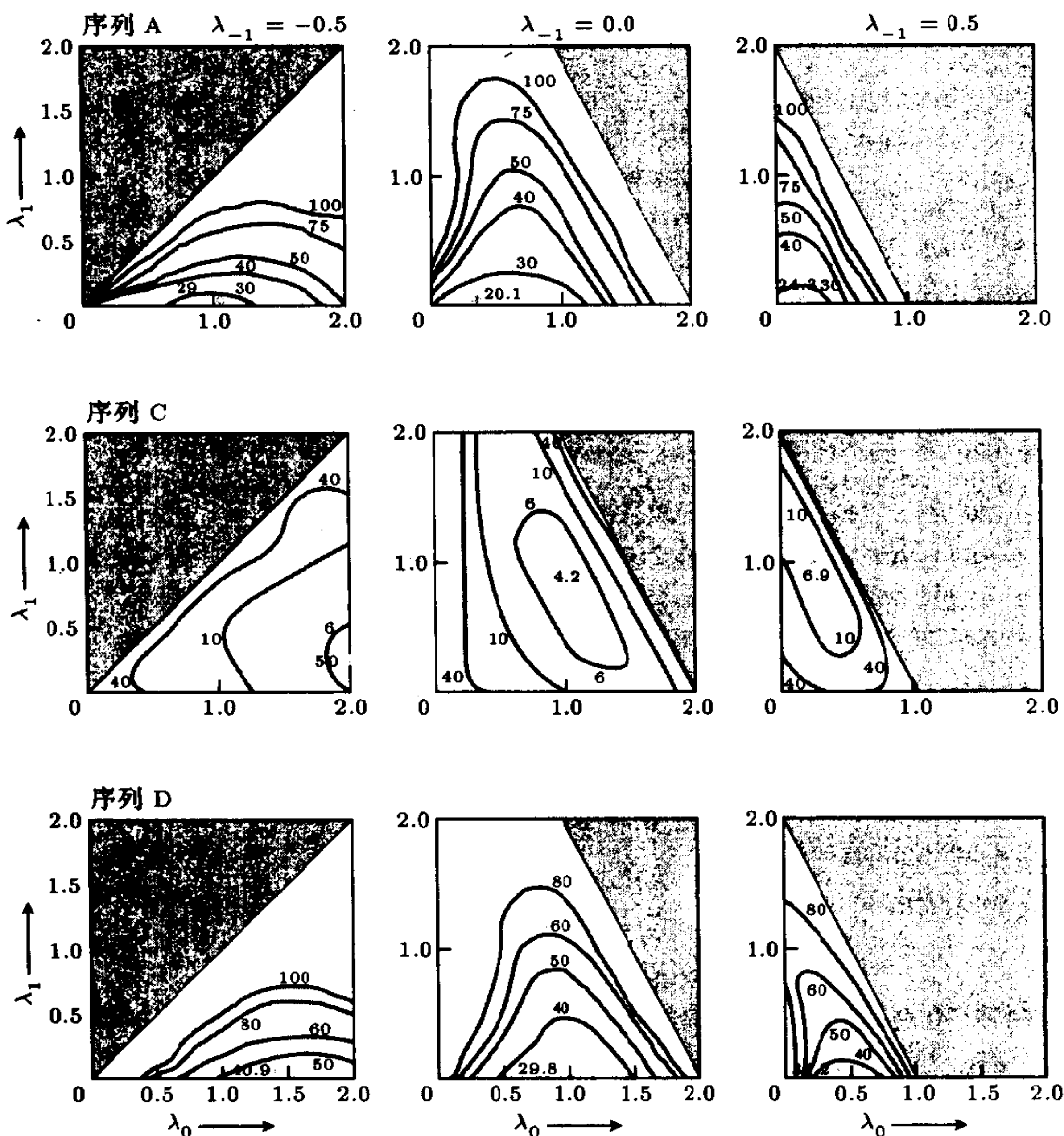


图 7.3 序列 A,C 和 D 的平方和等值线 (暗线标记可逆域的边界)

对每一种情形都是“精细化”下面的模型:

$$
\begin{aligned}
\nabla^2 z_t &= (1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)a_t \\
&= [1-(2-\lambda_{-1}-\lambda_0-\lambda_1)B-(\lambda_0+2\lambda_{-1}-1)B^2+\lambda_{-1}B^3]a_t
\end{aligned}
$$

或将模型表示为

$$
\nabla^2 z_t = (\lambda_{-1}\nabla^2+\lambda_0\nabla+\lambda_1)a_{t-1}+\nabla^2 a_t
$$

表 7.5 对序列 A,C 和 D 拟合 IMA 模型

序列	$\hat{\lambda}_{-1}$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	拟合序列
A	0	0.3	0	$\nabla z_t = 0.3a_{t-1} + \nabla a_t$
B	0	1.1	0.8	$\nabla^2 z_t = 1.1\nabla a_{t-1} + 0.8a_{t-1} + \nabla^2 a_t$
C	0	0.9	0	$\nabla z_t = 0.9a_{t-1} + \nabla a_t$

上面的模型已被拟合，在表 7.5 中列出了这种类型[①]模型中拟合最好的，这就是最后的结果。

在拟合过程中包含附加的参数 (特别是 λ_{-1}) 并不是完全必要的，我们引入它们是为了说明过拟合的效果，同时表明对于这些序列来说我们所识别的过程是多么贴切。

7.1.7 估计态势“状态良好”的描述；置信区域

对似然函数绘图当然不仅是为了指出最大似然值。似然函数的所有表现包含了来自数据的全部信息。在一些研究领域中可能发生这样的情形：似然函数有两个或更多的峰值 (例如可见 [39] 中的例子)，或是有分明的脊和尖峰。所有这些情形都有合乎逻辑的解释，在每一种情形，似然函数都试图告诉我们一些我们需要知道的东西。由此可知，若存在两个高度差不多相等的峰值，则意味着存在两组都可以解释数据的参数；若存在走向不正常的脊则意味着有一个参数与极大似然的参数值差别相当大，另外还伴有一个偏差较合理的参数，这样就能对数据作出解释了。这种特征确定了所谓的估计态势，为了理解估计态势，我们必须既从分析上同时又从图形上来考察似然函数。

解释似然函数需要小心 在解释似然函数时需要小心谨慎。例如在后面将要讨论的结果，其中假设了对数似然函数在最大值附近是近似为二次的，这显然不能适用于图 7.3 中所描述的三个参数的估计态势。然而，这些例子都是特殊的，因为我们特意对

[①]稍后在 7.2.5 节中我们将给出在某些情形下得出的略有改进的拟合，这是包含“平稳”自回归项且密切相关的一些模型。

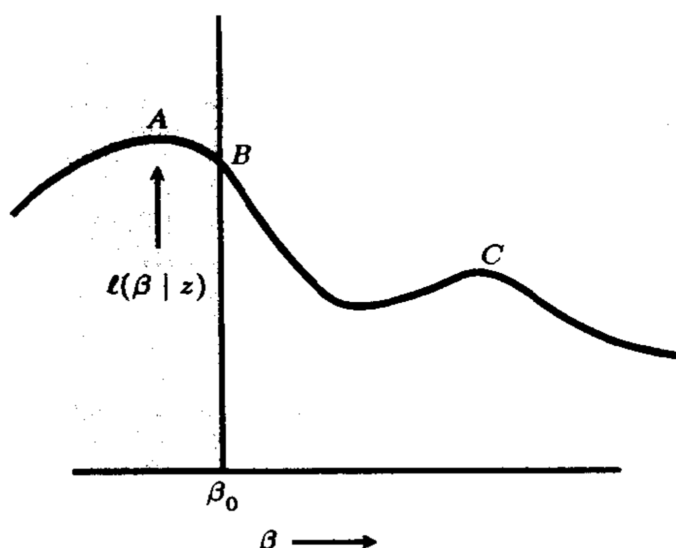


图 7.4 限定 $\beta>\beta_0$ 的假设似然函数

模型作过度拟合。如果较简单的模型是合理的，我们应该期望在高维参数空间的边界上接近最大值处做截面，由此得出极大似然的等值线。因此，如果所拟合的是简单的识别模型而不是超参数模型，则二次近似是可以使用的。

当似然函数位于或接近边界时，特别需要多加小心。考虑如图 7.4 所示情形，假设我们已知的前提是：参数 $\beta > \beta_0$ 。在 β 的允许域内极大似然值是 B(这里 $\beta = \beta_0$) 而不是 A 或 C 。应该注意到，本例中在最大值点处似然函数的一阶导数是非零的，因此二次近似显然不是似然函数的恰当描述。

在过去，能够提供处理似然函数的方法有些只是留下一些想象，而有些是不恰当的，这是由于实际应用者使用的方法不当而造成的错误。处理方法通常组成如下：

1. 对于对数似然函数微分并令一阶导数为零，从而得到极大似然 (ML) 估计。

2. 由对数似然函数的二阶导数或二阶导数的期望值得出这些估计值的近似方差和协方差。

机械地使用上述方法当然可能得出无意义的结果。之所以如此，首先是由于这样的基本事实 — 设定导数为零不一定能得到最大值；其次，似然函数所包含的信息能够用 ML 估计和对数似

然的二阶导数做充分的描述，当且仅当在所关注的区域上二次近似是恰当的。想要知道一个新的估计问题是否合乎上述要求，通常需要作仔细的分析和图形考察。

当我们开始考察一类估计问题时 (例如由 ARMA 模型的参数估计而引起的)，应该在较宽的范围画出似然函数，这是很重要的。在对某类具体模型的特性有了很好的理解，且对估计态势的了解表明一切都是稳妥可靠的，这时我们可以把当前考虑的区域做某种截取。在附录 A7.3 和 A7.4 中对这些结果将做更详细的描述。我们从考虑极大似然估计的方差和协方差的表达式入手，当对数似然函数近似为二次且样本量适中时，这种考虑是适合的。

为方便起见，下面定义一个向量 $\boldsymbol{\beta}$，它的 $k=p+q$ 个元素是自回归和滑动平均参数 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 。因此， ARMA 过程的全套 $p+q+1=k+1$ 个参数可记作 $\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a^2$; 或记作 $\boldsymbol{\beta},\sigma_a^2$, 或者就简单记作 $\boldsymbol{\xi}$ 。

ML 估计的方差与协方差　对于恰当参数化了的 ARMA 模型，在参数空间的相关区域[①]上，对数似然函数通常近似为 $\boldsymbol{\beta}$ (即 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$) 元素的二次形式，于是

$$l(\boldsymbol{\xi})=l(\boldsymbol{\beta},\sigma_a)\simeq l(\hat{\boldsymbol{\beta}},\sigma_a^2)+\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}l_{ij}(\beta_i-\hat{\beta}_i)(\beta_j-\hat{\beta}_j) \quad (7.1.18)$$

这里，在所考虑的近似下，导数

$$l_{ij}=\frac{\partial^2 l(\boldsymbol{\beta},\sigma_a)}{\partial\beta_i\partial\beta_j} \quad (7.1.19)$$

是常数。对较大的 n, 项 $f(\boldsymbol{\phi},\boldsymbol{\theta})$(或等价地，在 (7.1.5) 中的 $f(\boldsymbol{\beta})$) 的影响在多数情形下可以忽略。因此，如果 $S(\beta)$ 关于 $\boldsymbol{\beta}$ 是二次的，则 $l(\boldsymbol{\beta},\sigma_a)$ 也是二次的。另外，如果 (7.1.6) 中的条件期望 $[a_t|\boldsymbol{\beta},\boldsymbol{w}]$

[①] 称为 95% 置信域。

关于 $\boldsymbol{\beta}$ 的元素近似为局部线性的，则 $l(\boldsymbol{\beta},\sigma_a)$ 基本上是 $\boldsymbol{\beta}$ 的二次型.

对于中等和大样本情形，当局部二次近似 (7.1.18) 是合理的，便可以得出参数值的方差和协方差的有效近似式以及近似置信域.

参数 $\boldsymbol{\beta}$ 的信息阵　在 [91],[196] 中称 $k\times k$ 矩阵 $-E[I_{ij}]=\boldsymbol{i}(\boldsymbol{\beta})$ 为参数 $\boldsymbol{\beta}$ 的信息阵，这里的期望是对 $\boldsymbol{w}$ 的分布所取的。对于 σ_a 的给定值，在大样本情形， ML 估计 $\hat{\boldsymbol{\beta}}$ 的方差 - 协方差阵 $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ 由信息阵的逆阵给出，即

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}})\simeq\{-E[l_{ij}]\}^{-1} \tag{7.1.20}$$

例如，若 $k=2$, 大样本的方差 - 协方差阵为

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}})=\begin{bmatrix} V(\hat{\beta}_1) & \mathrm{cov}[\hat{\beta}_1,\hat{\beta}_2] \\ \mathrm{cov}[\hat{\beta}_1,\hat{\beta}_2] & V(\hat{\beta}_2)\end{bmatrix}\simeq-\begin{bmatrix} E[l_{11}] & E[l_{12}] \\ E[l_{12}] & E[l_{22}]\end{bmatrix}^{-1}$$

另外，从平稳过程 ARMA 得到的参数 $\hat{\boldsymbol{\beta}}$ 的 ML 估计被证明其渐近分布为多维正态，且具有均值向量 $\boldsymbol{\beta}$ 和协方差阵 $\boldsymbol{I}^{-1}(\boldsymbol{\beta})$ (如见 [142],[196],[215],[221]) 。

现在由 (7.1.5) 我们有

$$l_{ij}\simeq\frac{-S_{ij}}{2\sigma_a^2} \tag{7.1.21}$$

这里

$$S_{ij}=\frac{\partial^2 S(\boldsymbol{\beta}|\boldsymbol{w})}{\partial\beta_i\partial\beta_j}$$

进而，在大样本情形，我们用实际观测到的值来近似 l_{ij} 或 S_{ij} 的期望值，那么，由 (7.1.20) 我们可得

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}})\simeq\{-E[l_{ij}]\}^{-1}\simeq 2\sigma_a^2\{E[S_{ij}]\}^{-1}\simeq 2\sigma_a^2\{S_{ij}\}^{-1} \tag{7.1.22}$$

因此，对于 $k=2$ 有

$$V(\hat{\boldsymbol{\beta}}) \simeq 2\sigma_a^2 \begin{bmatrix} \dfrac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_1^2} & \dfrac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_2} \\ \dfrac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_2} & \dfrac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_2^2} \end{bmatrix}^{-1}$$

如果在参数空间的相关区域上 $S(\boldsymbol{\beta})$ 为 $\boldsymbol{\beta}$ 的精确二次形式，则所有的导数 S_{ij} 在这个区域上为常数。在实际中，S_{ij} 会有些变化，我们通常假设这些导数值在点 $\hat{\boldsymbol{\beta}}$ 或其附近被确定。在附录 A7.3 和 A7.4 中证明了 σ_a^2 的估计值。[①]

由下式给出

$$\hat{\sigma}_a^2 = \frac{S(\hat{\boldsymbol{\beta}})}{n} \tag{7.1.23}$$

并且对于大样本情形，$\hat{\sigma}_a^2$ 和 $\hat{\boldsymbol{\beta}}$ 是不相关的。最后，(7.1.22) 的元素可由下式估计

$$\operatorname{cov}[\hat{\beta}_i, \hat{\beta}_j] \simeq 2\hat{\sigma}_a^2 S^{ij} \tag{7.1.24}$$

这里，矩阵 $\{S^{ij}\}$ 由下式给出

$$\{S^{ij}\} = \{S_{ij}\}^{-1}$$

当 $j=i$ 时，表达式 (7.1.24) 理解为定义出了方差 $V(\hat{\beta}_i)$。

参数的近似置信域　特别地，这些结果使我们可以得到估计值的近似方差。对方差取平方根，我们就得出近似标准偏差，通常称为估计值的标准差。估计值 $\hat{\beta}_i$ 的标准差记为 $SE[\hat{\beta}_i]$。当我们必须同时考虑几个参数时，则需要一些办法来评价参数联合的精确性。一种实现的办法就是确定一个置信区间。可以证明 (见 [201])，一个 $1-\epsilon$ 的置信域的性质是：容量为 n 的重复样本设想

① 在 (7.1.23) 中，使用除数 $n-k=n-p-q$ 而不是 n，可以使结论得到改进，但是对于中等样本容量，这一修正只有微小差异。

为由同一总体中抽取，由每一个样本构造一个置信域，在这些置信域中包含参数真值点的所占比例为 $1-\epsilon$ 。

如果对于给定的 $\sigma_a^2, l(\boldsymbol{\beta}, \sigma_a^2)$ 在 $\hat{\boldsymbol{\beta}}$ 近似为 $\boldsymbol{\beta}$ 的二次型，则利用 (7.1.20)(又见附录 A7.1)，一个近似的 $1-\epsilon$ 置信域定义为

$$-\sum_{i,j} E[l_{ij}](\beta_i - \hat{\beta}_i)(\beta_j - \hat{\beta}_j) < \chi_\epsilon^2(k) \tag{7.1.25}$$

这里， $\chi_\epsilon^2(k)$ 是自由度为 k 的 χ^2 分布的临界值，超过该值的比例为 ϵ 。

另外，利用近似式 (7.1.22)，并用 (7.1.23) 的估计值代替 σ_a^2，则近似的置信域由下式给出

$$\sum_{i,j} S_{ij}(\beta_i - \hat{\beta}_i)(\beta_j - \hat{\beta}_j) < 2\sigma_a^2 \chi_\epsilon^2(k) \tag{7.1.26}$$

但是，对于二次曲面 $S(\boldsymbol{\beta})$ 有

$$S(\boldsymbol{\beta}) - S(\hat{\boldsymbol{\beta}}) = \frac{1}{2}\sum_{i,j} S_{i,j}(\beta_j - \hat{\beta}_j)(\beta_j - \hat{\beta}_j) \tag{7.1.27}$$

因此，由 (7.1.23) 和 (7.1.26)，我们最后得到这样的结果：近似的 $1-\epsilon$ 置信域由平方和曲面的等值线所界定，有

$$S(\boldsymbol{\beta}) = S(\hat{\boldsymbol{\beta}})\left[1 + \frac{\chi_\epsilon^2(k)}{n}\right] \tag{7.1.28}$$

计算近似置信区间和置信域的例子

1. 序列 B 。 对于序列 B, $S(\lambda)$ 及其差分的值在表 7.6 中列出。 $S(\lambda)$ 的二次差分不是常数，因此 $S(\lambda)$ 不是严格的二次型。

[①] 可以使用一种更好的近似，它基于 F 分布，并考虑了 $\hat{\sigma}_a^2$ 的样本分布．对于适度的样本这一修改对实际几乎没有差别．

然而，在 $\lambda = 0.85$ 到 $\lambda = 1.35$ 的范围内，$\nabla^2(S)$ 变化不大，因此，可以期望 (7.1.28) 能给出相当接近的近似。有了最小值 $S(\hat{\lambda}) = 19,216$，定义 95% 置信区间的临界值 $S(\lambda)$ 便可给出如下：

$$S(\lambda) - 19,216\left(1 + \frac{3.84}{368}\right) = 19,416$$

在图 7.1 中读出相应于 $S(\lambda) = 19,416$ 的 λ 值，我们得出近似置信区间为 $0.98 < \lambda < 1.19$.

另外，可以使用 (7.1.26)。用表 7.6 所给的 $\lambda = 1.1$ 处的二阶差分作为二阶导数的近似，我们得到

$$S_{11} = \frac{\partial^2 S}{\partial \lambda^2} \simeq \frac{390}{(0.1)^2}$$

用 (7.1.23) 还可得 $\hat{\sigma}_a = 19,216/368 = 52.5$。因此，由 (7.1.26) 定义的 95% 置信区间为

$$\frac{390}{(0.1)^2}(\lambda - 1.09)^2 < 2 \times 52.2 \times 3.84$$

即

$$|\lambda - 1.09| < 0.10$$

表 7.6　序列 B 对应于不同 λ 值的 $S(\lambda)$ 及其一阶和二阶差分

$\lambda=1-\theta$	$S(\lambda)$	$\nabla(S)$	$\nabla^2(S)$
1.5	23,928	2,333	960
1.4	21,595	1,373	634
1.3	20,222	739	476
1.2	19,483	263	406
1.1	19,220	−143	390
1.0	19,363	−533	422
0.9	19,896	−955	508
0.8	20,851	−1,463	691
0.7	22,314	−2,154	1069
0.6	24,468	−3,223	
0.5	27,691		

因此，置信区间为 $0.99 < \lambda < 1.19$, 这与前面的计算非常一致。

在本例中只有一个参数 λ, 用 (7.1.25) 和 (7.1.26) 等价于使用区间

$$\hat{\lambda} \pm u_{\epsilon/2}\hat{\sigma}(\hat{\lambda})$$

这里，$u_{\epsilon/2}$ 是偏差，它排除了正态分布的上部尾端比例为 $\epsilon/2$ 的部分。由 (7.1.24) 可得近似的标准差 $\hat{\sigma}(\hat{\lambda}) = \sqrt{2\hat{\sigma}_a^2 S_{11}^{-1}}$ 。在本例中

$$V(\hat{\lambda}) = 2\hat{\sigma}_a^2 S_{11}^{-1} = \frac{2 \times 52.2 \times 0.1^2}{390} = 0.00268$$

$\hat{\lambda}$ 的近似标准差为

$$\hat{\sigma}(\hat{\lambda}) = \sqrt{\text{var}[\hat{\lambda}]} = 0.052$$

因此，与前一样，近似的置信区间是

$$\hat{\lambda} \pm 1.96\hat{\sigma}(\hat{\lambda}) = 1.09 \pm 0.10$$

最后，我们在稍后的 7.2 节中还要说明解析地计算 (7.1.20) 的可能性。对来自于 MA(1) 过程的大样本得到

$$V(\hat{\lambda}) \simeq \frac{\lambda(2-\lambda)}{n}$$

对于本例，用 $\hat{\lambda} = 1.09$ 代替 λ, 我们求出

$$V(\hat{\lambda}) \simeq 0.00269$$

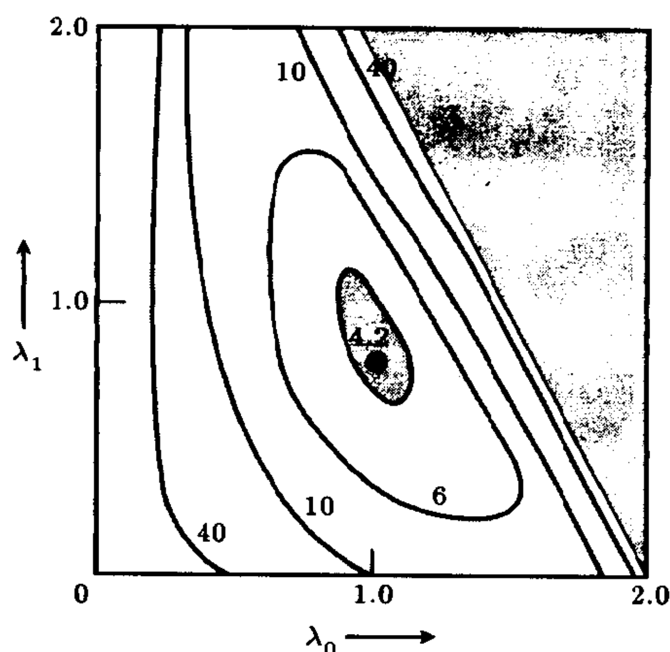


图 7.5 对于序列 C 假设其模型为 (0,2,2) 阶，平方和等值线内 (阴影)95% 的近似置信域

这与先前的估计非常一致，从而得出同样的标准差 0.052 和同样的置信区间。

2. 序列 C 。　在识别序列 C 时，所得的一个模型是 (0,2,2) 过程. 为了说明 (7.1.28) 对于多于一个参数的应用，图 7.5 对序列 C 的 λ_0 和 λ_1 给出了近似的 95% 置信域 (阴影部分)．在本例中， $S(\hat{\lambda}) = 4.20, n = 224$, 及 $\chi^2_{0.05}(2) = 5.99$, 因此，近似的 95% 置信域被等值线所界定，相应的值为

$$S(\lambda_0, \lambda_1) = 4.20\left(1 + \frac{5.99}{224}\right) = 4.31$$

7.2 非线性估计

7.2.1 一般的近似方法

在研究新的估计问题时，平方和函数作图是特别重要的，因为它保证了估计态势中任何细节都会显现出来。当我们认为没有出现不正常现象时，就可以使用其它的方法。

我们已经知道，对于极大多数场合，极大似然估计可以用最小二乘估计很好地近似，即是使下式达到最小

$$S(\boldsymbol{\phi},\boldsymbol{\theta})=\sum_{t=1}^{n}[a_t]^2+[\boldsymbol{e}_*]'\boldsymbol{\Omega}^{-1}[\boldsymbol{e}_*]$$

在实际中，可用反向模型方法由有限和 $\sum_{t=1-Q}^{n}[a_t]^2$ 来近似。

通常，在使平方和 $\sum_{t=1}^{n}[f_t(\boldsymbol{\beta})]^2$ 关于 $\boldsymbol{\beta}$ 极小化时，如果每个 $f_t(\boldsymbol{\beta})(t=1,2,\cdots,n)$ 都是参数 $\boldsymbol{\beta}$ 的线性函数，那么，问题会变得相当简单。现在我们来说明：$[a_t]$ 的线性性质相对于自回归参数 $\boldsymbol{\phi}$ 以及相对于滑动平均参数 $\boldsymbol{\theta}$ 是有所不同的。

对于纯自回归过程，$[a_t]=\phi(B)[\widetilde{w}_t]$ 及

$$\frac{\partial[a_t]}{\partial\phi_i}=-[\widetilde{w}_{t-i}]+\phi(B)\frac{\partial[\widetilde{w}_t]}{\partial\phi_i}$$

现在对 $u>0,[\widetilde{w}_u]=\widetilde{w}_u$ 且 $\partial[\widetilde{w}_u]/\partial\phi_i=0$，而对 $u<0,[\widetilde{w}_u]$ 和 $\partial[\widetilde{w}_u]/\partial\phi_i$ 都是 $\boldsymbol{\phi}$ 的函数。因此，除了“初值”的作用，$[a_t]$ 关于 $\boldsymbol{\phi}$ 是线性的。与此不同，对于纯滑动平均过程有

$$[a_t]=\theta^{-1}(B)[\widetilde{w}_t]\ ,\quad \frac{\partial[a_t]}{\partial\theta_j}=\theta^{-2}(B)[\widetilde{w}_{t-j}]+\theta^{-1}(B)\frac{\partial[\widetilde{w}_t]}{\partial\theta_j}$$

于是 $[a_t]$ 总是参数的非线性函数。

我们在 7.3 节将会看到，为了得到最小二乘估计和极大似然估计，对于自回归过程是特别简单的。在本节中我们将说明如何运用线性最小二乘迭代，得出任何 ARMA 模型的参数估计值。

模型的线性化　在下文中，我们继续使用 $\boldsymbol{\beta}$ 作为 $k=p+q$ 个参数 $(\boldsymbol{\phi},\boldsymbol{\theta})$ 的一般记号。于是，我们要极小化

$$S(\boldsymbol{\phi},\boldsymbol{\theta})\simeq\sum_{t=1-Q}^{n}[a_t|\widetilde{\boldsymbol{w}},\boldsymbol{\beta}]^2=\sum_{t=1-Q}^{n}[a_t]^2$$

将 $[a_t]$ 用 Taylor 级数展开，在该值附近有一组推测的对应参数值 $\boldsymbol{\beta}_0' = (\beta_{1,0}, \beta_{2,0}, \cdots, \beta_{k,0})$ 。我们近似有

$$[a_t] = [a_{t,0}] - \sum_{i=1}^{k} (\beta_i - \beta_{i,0}) x_{t,i} \tag{7.2.1}$$

其中

$$[a_{t,0}] = [a_t | \boldsymbol{w}, \boldsymbol{\beta}_0]$$

及

$$x_{t,i} = -\left.\frac{\partial [a_t]}{\partial \beta_i}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

现在，若 $\boldsymbol{X}$ 记 $(n+Q) \times k$ 矩阵 $\{x_{t,i}\}$，则 (7.2.1) 的 $n+Q$ 个方程可以表示为

$$[\boldsymbol{a}_0] = \boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + [\boldsymbol{a}]$$

这里， $[\boldsymbol{a}_0]$ 和 $[\boldsymbol{a}]$ 是有 $n+Q$ 个元素的列向量。

调整量 $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ 使得 $S(\boldsymbol{\beta}) = S(\boldsymbol{\phi}, \boldsymbol{\theta}) = [\boldsymbol{a}]'[\boldsymbol{a}]$, 达到极小，我们可以通过线性最小二乘求得，也就是通过用 $[\boldsymbol{q}_0]$ 在 $\boldsymbol{X}$ 上作“回归”得到，由于 $[a_t]$ 关于参数 $\boldsymbol{\beta}$ 不是严格线性的，因此仅仅一次调整不会立刻得出最小平方值。把调整量代入得出调整后的值，以此作为新的推测值，重复该过程直到收敛。如果一开始用了好的推测值，例如在模型识别阶段得到的那些估计值，则收敛会较快，但如果用了相当坏的推测值，那么，迭代过程可能根本不收敛。

7.2.2 导数的数值估计

正如我们后文所说明的，导数 $x_{t,i}$ 可以直接得出。然而，对于机器计算来说，人们发现通用的非线性最小二乘程序非常令人满意，而其中的导数是用数值方法得出的。具体做法是通过“每次一个 ”的办法使参数变动。因此，对于给定的模型，值 $[a_t | \boldsymbol{w}, \beta 1, 0, \beta_{2,0}, \cdots, \beta_{k,o}]$ 对于 $t = 1 - Q, \cdots, n$ 是用递推方

法计算的，无论如何，使用“反向预报值”作为初值或许还是需要的。然后对 $[a_t|\boldsymbol{w},\beta_{1,0}+\delta_1,\beta_{2,0},\cdots,\beta_{k,0}]$ 重复递推计算，继而又对 $[a_t|\boldsymbol{w},\beta_{1,0},\beta_{2,0}+\delta_2,\cdots,\beta_{k,0}]$ 计算，等等。由

$$x_{t,i}=\frac{[a_t|\boldsymbol{w},\beta_{1,0},\cdots,\beta_{i,0},\cdots,\beta_{k,0}]-[a_t|\boldsymbol{w},\beta_{1,0},\cdots,\beta_{i,0}+\delta_i,\cdots,\beta_{k,0}]}{\delta_i} \tag{7.2.2}$$

给出了具有足够精度的所需导数负值。上述获得导数的数值方法具有广泛适用的优点，它只需要我们编程计算 $[a_t]$，而不是计算它们的导数。通用的非线性估计程序基本上只要求输入如何计算 $[a_t]$，这些程序是广泛适用的 [33] 。在一些版本中需要预先选择 δ，另一些版本则由程序本身通过预先的迭代得出适合的 δ 。一些程序还包括一些殊功能，可以避免漏掉最大值并加速收敛 [143] 。

如果得出的最小二乘解不在或不接近约束的边界，那么，来自最后一次迭代的 $\boldsymbol{X}=\boldsymbol{X}_{\hat{\boldsymbol{\beta}}}$ 的值可以用来计算近似的方差、协方差和置信区间。因此，$(\boldsymbol{X}'_{\beta}\boldsymbol{X}_{\beta})^{-1}\sigma_a^2$ 近似于 $\hat{\boldsymbol{\beta}}$ 的方差－协方差阵，σ_a^2 可由 $S(\hat{\boldsymbol{\beta}})/n$ 估计。

对 (0,1,1) 过程的应用　作为简单的说明实例，考虑对序列 A 拟合 (0,1,1) 过程：

$$w_t=\nabla z_t=(1-\theta B)a_t$$

其中 $\mu=E[w_t]=0$ 。表 7.7 列出了对推测值 $\theta_0=0.5$ 进行计算的开始一段。$[a_0]$ 的反向预报值实际上是通过设 $[e_7]=0$ 并利用反向递推 $[e_t]=[w_t]+\theta[e_{t+1}]$ 而得到的。在序列更远处开始反向递推可取得更好的精度。对于 $\theta=0.50$ 和 $\theta=0.51$ 顺次使用 $[a_t]=\theta[a_{t-1}]+[w_t]$，求得 $[a_t]$ 相应的值列在表中的第四列和第五列，在表中一起列出的 x_t 是由 (7.2.2) 得到的负导数值。为了求出 θ 的最初调整值，我们计算

$$\theta-\theta_0=\frac{\sum\limits_{t=0}^{n}[a_{t,0}]x_t}{\sum\limits_{t=0}^{n}x_t^2}$$

表 7.7 对于序列 A 的数据用数值方法计算导数的说明

t	z_t	w_t	$[a_{t,0}]=[a_t\|0.50]$	$[a_t\|0.51]$	$x_t=100([a_t\|0.50]$ $-[a_t\|0.51])$
0	17.0		0.2453	0.2496	−0.43
1	16.6	−0.40	−0.2773	−0.2727	−0.46
2	16.3	−0.30	−0.4783	−0.4391	0.04
3	16.1	−0.20	−0.4193	−0.4239	0.46
4	17.1	1.00	0.7903	0.7838	0.65
5	16.9	−0.20	0.1952	0.1997	−0.45
6	16.8	−0.10	−0.0024	0.0019	−0.43
7	17.4	0.60	0.5988	0.6010	−0.22
8	17.1	−0.30	−0.0006	0.0065	−0.71
9	17.0	−0.10	−0.1003	−0.0967	−0.36
10	16.7	−0.30	−0.3502	−0.3493	−0.09

在这个例子中，使用整个序列的 197 个观测值，经过四次迭代后可收敛。计算过程如下：

迭代次数	0	1	2	3	4	5
θ	0.50	0.63	0.68	0.69	0.70	0.70

对于任何要求的精度，用这种方法一般都可以得到使 $S(\boldsymbol{\phi},\boldsymbol{\theta})$ 最小的 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 值。该方法是颇具吸引力的，因为它的使用不需要编制求导程序，这一点不同于 $[a_t]$ 的计算。另外对于最终的计算效果不必采取特别的处理。

我们现在来说明，直接求得导数也是可能的，但是需要做另外的递推计算。

7.2.3 导数的直接估算

为了说明这种方法，考虑一个 ARMA(1,1) 模型就够了，该模型可记为下面任一种形式

$$e_t = w_t - \phi w_{t+1} + \theta e_{t+1}$$
$$a_t = w_t - \phi w_{t-1} + \theta a_{t-1}$$

我们在 7.1.4 节中已说明，该模型的两种形式可以交替使用，其中

一个给出开始递推需要的初值，而另一个用来递推。我们假设已经进行了第一次计算，得到值 $[e_t]$ 和 $[a_t]$，及 $[w_0],[w_{-1}],\cdots,[w_{1-Q}]$。和 7.1.5 节一样，$[w_{-Q}],[w_{-Q-1}],\cdots$ 的值，从而 $[a_{-Q}],[a_{-Q-1}],\cdots$ 的值都可以忽略。我们现在来说明，在计算导数时可以使用类似的对偶计算。

用 $a_t^{(\phi)}$ 记偏导数 $\partial[a_t]/\partial\phi$，我们得到

$$e_t^{(\phi)} = w_t^{(\phi)} - \phi w_{t+1}^{(\phi)} + \theta e_{t+1}^{(\phi)} - [w_{t+1}] \tag{7.2.3}$$

$$a_t^{(\phi)} = w_t^{(\phi)} - \phi w_{t-1}^{(\phi)} + \theta a_{t-1}^{(\phi)} - [w_{t-1}] \tag{7.2.4}$$

$$e_t^{(\theta)} = w_t^{(\theta)} - \phi w_{t+1}^{(\theta)} + \theta e_{t+1}^{(\theta)} - [e_{t+1}] \tag{7.2.5}$$

$$a_t^{(\theta)} = w_t^{(\theta)} - \phi w_{t-1}^{(\theta)} + \theta a_{t-1}^{(\theta)} - [a_{t-1}] \tag{7.2.6}$$

现在

$$\left.\begin{array}{l}[w_t] = w_t \\ w_t^{(\phi)} = w_t^{(\theta)} = 0\end{array}\right\} \quad (t = 1,2,\cdots,n) \tag{7.2.7}$$

和

$$[e_{-j}] = 0 \qquad (j = 0,1,\cdots,n) \tag{7.2.8}$$

考虑方程 (7.2.3) 和 (7.2.4)。通过在 (7.2.3) 中令 $e_{n+1}^{(\phi)} = 0$，我们可以着手利用 (7.2.7) 和 (7.2.8) 的反向递推，最终我们可以计算出 $w_{-j}^{(\phi)}, j = 0,1,\cdots,Q-1$。由于可设 $a_{-Q}^{(\phi)}, a_{-Q-1}^{(\phi)},\cdots$ 为零，我们现在可以利用 (7.2.4) 递推计算需要的导数 $a_t^{(\phi)}$。与此类似，可用 (7.2.5) 和 (7.2.6) 来计算导数 $a_t^{(\theta)}$。

为了说明起见，仍考虑序列 A 中前一部分数据有关 $x_t = -\partial[a_t]/\partial\theta$ 的计算，在表 7.7 中完全是用数值方法进行计算的. 表 7.8 列出了采用下面式子的相应计算:

$$e_t^{(\theta)} = w_t^{(\theta)} + \theta e_{t+1}^{(\theta)} + [e_{t+1}]$$

$$-x_t = a_t^{(\theta)} = w_t^{(\theta)} + \theta a_{t-1}^{(\theta)} + [a_{t-1}]$$

已算出的 $[a_t]$ 和 $[e_t]$ 值首先列入表中，示例中 $e_t^{(\theta)}$ 的计算是从令 $e_6^{(0)} = 0$ 开始。可以看出，x_t 值非常接近于表 7.7 中列出的用完全数值方法得到的值。

表 7.8 对于序列的数据递推计算导数的说明

t	$[a_{t-1}]$	$\theta a_{t-1}^{(\theta)}$	$x_t=-a_t^{(\theta)}$	$e_t^{(\theta)}$	$\theta e_{t+1}^{(\theta)}$	$[e_{t+1}]$
0			−0.43	$(-w_0^{(\theta)}=-0.43)$	0.06	−0.49
1	0.25	0.22	−0.47	0.12	0.30	−0.18
2	−0.28	0.24	0.04	0.60	0.36	0.24
3	−0.44	−0.02	0.46	0.73	−0.15	0.88
4	−0.42	−0.23	0.65	−0.30	−0.05	−0.25
5	0.79	−0.33	−0.46	−0.10	0	−0.10
6	0.20	0.23	−0.43	0		−0.10
7	0.00	0.22	−0.22			
8	0.60	0.11	−0.71			
9	0.00	0.35	−0.35			
10	−0.10	−0.18	−0.08			

7.2.4 条件模型的一般最小二乘算法

我们有时对长序列使用一种近似方法：设 a 的初值为它们的非条件期望 — 零，从而 x 的初值也为零，然后直接用正向递推继续进行。因此对先前的例子我们可应用方程

$$a_t = \theta a_{t-1} + w_t \quad , \quad x_t = -a_t^{(\theta)} = \theta x_{t-1} - a_{t-1}$$

其效果是在 a_t 和 x_t 中都引入了瞬态，后者衰减较慢，这是因为 x 依赖于 a 。作为说明，该方法用于当前的例子而得到的 a 和 x 值可以计算出来。我们发现，虽然一开始并不一致，但是，a 从 $t=4$ 以后，在两位小数上是一致的，而 x 从 $t=8$ 后一致。在一些情形，若有充足的数据 (如 200 或更多的观测值)，这种近似的影响可以消除，比如这时可以放弃一些 (如前 10 个) 计算值，这样做是以信息损失为代价的。

如果我们采用这种近似，可以得到对于条件模型的一种有意义的一般算法。一般模型可以写为

$$a_t = \theta^{-1}(B)\phi(B)\widetilde{w}_t$$

这里，$w_t = \nabla^d z_t, \widetilde{w}_t = w_t - \mu$ 及

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_i B^i - \cdots - \theta_q B^q$$
$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_j B^j - \cdots - \phi_p B^p$$

如果参数 $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$ 的第一个推测值是 $\boldsymbol{\beta} = (\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$, 那么

$$a_{t,0} = \theta_0^{-1}(B)\phi_0(B)\widetilde{w}_t$$

及

$$-\left.\frac{\partial a_t}{\partial \phi_j}\right|_{\boldsymbol{\beta}_0} = u_{t,j} = u_{t-j} \quad , \qquad -\left.\frac{\partial a_t}{\partial \theta_i}\right|_{\boldsymbol{\beta}_0} = v_{t,i} = v_{t-i}$$

这里

$$u_t = \theta_0^{-1}(B)\widetilde{w}_t \qquad = \phi_0^{-1}(B)a_{t,0} \tag{7.2.9}$$
$$v_t = -\theta_0^{-2}(B)\phi_0(B)\widetilde{w}_t = -\theta_0^{-1}(B)a_{t,0} \tag{7.2.10}$$

a, u 及 v 可递推计算，设 a, u 和 v 的初始值为 0, 递推公式如下：

$$a_{t,0} = \widetilde{w}_t - \phi_{1,0}\widetilde{w}_{t-1} - \cdots - \phi_{p,0}\widetilde{w}_{t-p} + \theta_{1,0}a_{t-1,0} + \cdots + \theta_{q,0}a_{t-q,0} \tag{7.2.11}$$
$$u_t = \theta_{1,0}u_{t-1} + \cdots + \theta_{q,0}u_{t-q} + \widetilde{w}_t \tag{7.2.12}$$
$$= \phi_{1,0}u_{t-1} + \cdots + \phi_{p,0}u_{t-p} + a_{t,0} \tag{7.2.13}$$
$$v_t = \theta_{1,0}v_{t-1} + \cdots + \theta_{q,0}u_{t-q} - a_{t,0} \tag{7.2.14}$$

与 (7.2.1) 对应的近似线性回归方程成为

$$a_{t,0} = \sum_{j=1}^{p}(\phi_j - \phi_{j,0})u_{t-j} + \sum_{i=1}^{q}(\theta_i - \theta_{i,0})v_{t-i} + a_t \tag{7.2.15}$$

于是调整值是 $a_{t,0}$ 关于 u_{t-j} 和 v_{t-i} 的回归系数。通过把第一次推测值 $(\boldsymbol{\phi}_0, \boldsymbol{\theta}_0)$ 加上调整值就构成了一组“第二次”推测值，在

第二次迭代中用它代替 $(\boldsymbol{\phi}_0,\boldsymbol{\theta}_0)$，并计算 $a_{t,0}u_t$ 和 v_t 的新值，如此直至最终收敛。

算法的另一种形式　近似线性展开 (7.2.15) 可记为如下形式

$$
\begin{aligned}
a_{t,0} &= \sum_{j=1}^{p}(\phi_j-\phi_{j,0})B^j\phi_0^{-1}(B)a_{t,0}-\sum_{i=1}^{q}(\theta_i-\theta_{i,0})B^i\theta_0^{-1}(B)a_{t,0}+a_t \\
&= -[\phi(B)-\phi_0(B)]\phi_0^{-1}(B)a_{t,0}+[\theta(B)-\theta_0(B)]\theta_0^{-1}(B)a_{t,0}+a_t
\end{aligned}
$$

即

$$
a_{t,0} = -\phi(B)[\phi_0^{-1}(B)a_{t,0}]+\theta(B)[\theta_0^{-1}(B)a_{t,0}]+a_t \tag{7.2.16}
$$

该式用一种有趣的形式给出了算法

对 IMA(0,2,2) 过程的应用　为了说明条件近似的计算，我们考虑对序列 C 的最小二乘估计 $\hat{\theta}_1,\hat{\theta}_2$，使用的 (0,2,2) 模型是：

$$
w_t = (1-\theta_1B-\theta_2B^2)a_t
$$

其中

$$
\begin{aligned}
w_t &= \nabla^2 z_t \\
a_{t,0} &= w_t+\theta_{t,0}a_{t-1,0}+\theta_{2,0}a_{t-2,0} \\
v_t &= -a_{t,0}+\theta_{1,0}v_{t-1}+\theta_{2,0}v_{t-2}
\end{aligned}
$$

表 7.9　对 IMA(0,2,2) 过程参数 θ_1 和 θ_2 的非线性估计

t	z_t	∇z_t	$\nabla^2 z_t=w_t$	$a_{t,0}$	v_{t-1}	v_{t-2}
1	26.6			0	0	0
2	27.0	0.4		0	0	0
3	27.1	0.1	−0.3	−0.300	0	0
4	27.2	0.1	0.0	−0.030	0.300	0
5	27.3	0.1	0.0	−0.033	0.060	0.300
6	26.9	−0.4	−0.5	−0.533	0.069	0.060
7	26.4	−0.5	−0.1	−0.156	0.546	0.069
8	26.0	−0.4	0.1	0.039	0.218	0.546
9	25.8	−0.2	0.2	0.189	0.038	0.218

表 7.10 θ_1 和 θ_2 迭代的收敛性

迭代次数	θ_1	θ_2
0	0.1000	0.1000
1	0.1247	0.1055
2	0.1266	0.1126
3	0.1286	0.1141
4	0.1290	0.1149
5	0.1292	0.1151
6	0.1293	0.1152
7	0.1293	0.1153
7	0.1293	0.1153

表 7.9 列出了初值 $\theta_{1,0}=0.1$ 和 $\theta_{2,0}=0.1$ 的计算。

通过 $a_{t,0}$ 对于 v_{t-1} 和 u_{t-2} 的"回归"可以得到 $\theta_{1,0}$ 和 $\theta_{2,0}$ 的第一次调整值，重复该过程直到收敛。迭代过程如表 7.10 所示，使用的初值为 $\theta_{1,0}=0.1$ 和 $\theta_{2,0}=0.1$ 。

7.2.5 对于序列 A 到 F 拟合模型的总结

在表 7.11 中我们总结了对序列 A 到 F 用 7.2.1 和 7.2.2 节的迭代最小二乘法所拟合的模型。这些拟合的模型是在第 6 章中识别的，有关内容总结在表 6.4 中。我们从表 7.11 中看到，对于序列 A,C 和 D 识别并随之拟合了两种可能的模型。对于序列 A 和 D, 另一种模型涉及到用平稳自回归算子 $(1-\theta B)$ 。代替非平稳算子 $(1-B)$ 考察表 7.11 可知，在这两个例子中，自回归模型都得出较小的残差方差，尽管正如已指出的那样，两个模型非常相似，即使用平稳模型可能拟合得稍好，但在这两个例子中可能还应选择 (0,1,1) 模型，因为它不象平稳模型那样要假设序列有一个固定的均值。这一点在预测序列的未来值时特别重要，因为如果水平确实改变了，那么，$d>0$ 的模型将继续跟踪变化，而 $d=0$ 的模型则停滞在已经过时的均值水平上。

如同表 7.2.1 中所述，表 7.11 中系数下面的界限表示由协方差阵 $(\boldsymbol{X}'_{\hat{\boldsymbol{\beta}}}\boldsymbol{X}_{\hat{\boldsymbol{\beta}}})^{-1}\sigma_a^2$ 所得到的估计值的标准差。注意，对太阳黑子

表 7.11 对于序列 A 到 F 拟合模型的总结

序列	观测值个数	拟合模型	残差方差
A	197	$z_t-0.92z_{t-1}=1.45+a_t-0.58a_{t-1}$ (±0.04) (±0.08)	0.097
		$\nabla z_t=a_t-0.70a_{t-1}$ (±0.05)	0.101
B	369	$\nabla z_t=a_t-0.09a_{t-1}$ (±0.05)	52.2
C	226	$\nabla z_t-0.82\nabla z_{t-1}=a_t$ (±0.04)	0.018
		$\nabla z_t=a_t-0.13a_{t-1}-0.12a_{t-2}$ (±0.07) (±0.07)	0.019
D	310	$\nabla z_t-0.87\nabla z_{t-1}=1.17+a_t$ (±0.03)	0.090
		$\nabla z_t=a_t-0.06a_{t-1}$ (±0.06)	0.096
E	100	$z_t=14.35+1.42z_{t-1}-0.73z_{t-2}+a_t$ (±0.07) (±0.07)	228
		$\nabla z_t=11.31+1.57z_{t-1}-1.02z_{t-2}+0.21z_{t-3}+a_t$ (±0.15) (±0.10)	218
F	70	$\nabla z_t=58.87-0.34z_{t-1}+0.19z_{t-2}+a_t$ (±0.12) (±0.12)	113

* 在每个估计值下面带有 (±) 的值表示该估计值的标准差

** 由 $S(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})/n$ 得出

序列 E 拟合的 AR(3) 中，估计值 $\hat{\phi}_3$ 是其标准差的 2.1 倍，这表明与二阶自回归过程相比较，三阶自回归过程得到在一定程度上更好的模型。这与 Moran[146] 得到的结论一致。

7.2.6 大样本信息阵和协方差估计

我们用 $\boldsymbol{X}=[\boldsymbol{U}:\boldsymbol{V}]$ 记 $n\times(p+q)$ 阵，其中带有时间滞后的 u 和 v 在 (7.2.13) 和 (7.2.14) 中定义，假设 $\boldsymbol{\beta}_0$ 中的元素为参数真值，并有足够大的样本容量 n，以使端值效应可以忽略。那么，对

于混合 ARMA 模型 $(\boldsymbol{\phi},\boldsymbol{\theta})$ 的信息阵是

$$\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})=E\begin{bmatrix}\boldsymbol{U}'\boldsymbol{U} & \boldsymbol{U}'\boldsymbol{V}\\ \boldsymbol{V}'\boldsymbol{U} & \boldsymbol{V}'\boldsymbol{V}\end{bmatrix}\sigma_a^{-2}=E(\boldsymbol{X}'\boldsymbol{X})\sigma_a^{-2} \tag{7.2.17}$$

即

$$=n\sigma_a^{-2}\left[\begin{array}{cccc|cccc}
\gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(p-1) & \gamma_{uv}(0) & \gamma_{uv}(-1) & \cdots & \gamma_{uv}(1-q)\\
\gamma_{uu}(1) & \gamma_{uu}(0) & \cdots & \gamma_{uu}(p-2) & \gamma_{uv}(1) & \gamma_{uv}(0) & \cdots & \gamma_{uv}(2-q)\\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots\\
\gamma_{uu}(p-1) & \gamma_{uu}(p-2) & \cdots & \gamma_{uu}(0) & \gamma_{uv}(p-1) & \gamma_{uv}(p-2) & \cdots & \gamma_{uv}(p-q)\\
\hline
\gamma_{uv}(0) & \gamma_{uv}(1) & \cdots & \gamma_{uv}(p-1) & \gamma_{vv}(0) & \gamma_{vv}(1) & \cdots & \gamma_{vv}(q-1)\\
\gamma_{uv}(-1) & \gamma_{uv}(0) & \cdots & \gamma_{uv}(p-2) & \gamma_{vv}(1) & \gamma_{vv}(0) & \cdots & \gamma_{vv}(q-2)\\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots\\
\gamma_{uv}(1-q) & \gamma_{uv}(2-q) & \cdots & \gamma_{uv}(p-q) & \gamma_{vv}(q-1) & \gamma_{vv}(q-2) & \cdots & \gamma_{vv}(0)
\end{array}\right] \tag{7.2.18}$$

这里，$\gamma_{uu}(k)$ 和 $\gamma_{vv}(k)$ 是 u 和 v 的自协方差，$\gamma_{uv}(k)$ 是互协方差，定义为

$$\gamma_{uv}(k)=\gamma_{vu}(-k)=E[u_t v_{t+k}]=E[v_t u_{t-k}]$$

极大似然估计的大样本协方差阵可由下式得出

$$\boldsymbol{V}(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})\simeq\boldsymbol{I}^{-1}(\boldsymbol{\phi},\boldsymbol{\theta})$$

$\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})$ 以及 $\boldsymbol{V}(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})$ 的值可以这样得到：在 $\boldsymbol{\beta}_0=\hat{\boldsymbol{\beta}}$ 处求 u 和 v 的值，并在 (7.2.17) 中去掉期望符号，于是得出

$$\hat{\boldsymbol{V}}(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})=(\boldsymbol{X}_{\hat{\boldsymbol{\beta}}}'\boldsymbol{X}_{\hat{\boldsymbol{\beta}}})^{-1}\sigma_a^2$$

或是将自协方差和互协方差的标准样本估计代入 (7.2.18) 中。为了得到大样本的理论结果，我们注意到，$\boldsymbol{\beta}_0$ 的元素等于真值时，

方程 (7.2.13) 和 (7.2.14) 意味着导数序列 u_t 和 v_t 满足如下定义的自回归过程:

$$\phi(B)u_t = a_t \quad , \quad \theta(B)v_t = -a_t$$

所以，在 (7.2.18) 中出现的自协方差来自于纯自回归过程，而互协方差是由 a_t 生成的这样两个过程之间互协方差的负值。

我们通过几个例子来说明这一结果的应用。

AR(p) 和 MA(q) 过程参数估计值的协方差阵　设 $\boldsymbol{\Gamma}_p(\boldsymbol{\phi})$ 是来自一个 AR(p) 过程的 p 个顺序观测值生成的 $p\times p$ 自协方差阵，AR 过程的相应参数为 $\boldsymbol{\phi}' = (\phi_1, \phi_2, \cdots, \phi_p)$ 。于是利用 (7.2.18), 可给出估计值 $\hat{\boldsymbol{\phi}}$ 的 $p\times p$ 协方差阵为

$$\boldsymbol{V}(\hat{\boldsymbol{\phi}}) \simeq n^{-1}\sigma_a^2\boldsymbol{\Gamma}_p^{-1}(\boldsymbol{\phi}) \tag{7.2.19}$$

设 $\boldsymbol{\Gamma}_q(\boldsymbol{\theta})$ 是来自一个 AR(q) 过程的 q 个顺序观测值生成的 $q\times q$ 自协方差阵， AR 过程的相应参数为 $\boldsymbol{\theta}' = (\theta_1, \theta_2, \cdots, \theta_q)$ 。于是利用 (7.2.18), 可给出 MA(q) 模型的估计值 $\hat{\boldsymbol{\theta}}$ 的 $q\times q$ 协方差阵为

$$\boldsymbol{V}(\hat{\boldsymbol{\theta}}) \simeq n^{-1}\sigma_a^2\boldsymbol{\Gamma}_q^{-1}(\boldsymbol{\theta}) \tag{7.2.20}$$

有时用 $\phi(B)$ 和 $\theta(B)$ 的零点来参数化 ARMA 过程是很有用的。在这种情形下可以得到参数估计值协方差阵的一种特别简单的形式。

* ARMA 过程零点的协方差　考虑 ARMA(p,q) 过程按其零点 (假设为实值) 的参数化，即

$$\prod_{i=1}^{p}(1-G_iB)\widetilde{w}_t = \prod_{j=1}^{q}(1-H_jB)a_t$$

* 从此至 7.3 之前的内容在初次阅读时可以略去。

或

$$a_t = \prod_{i=1}^{p}(1-G_iB)\prod_{j=1}^{q}(1-H_jB)^{-1}\widetilde{w}_t$$

于是 a 的导数为

$$u_{t,i} = -\frac{\partial a_t}{\partial G_i} = (1-G_iB)^{-1}a_{t-1}$$
$$v_{t,j} = -\frac{\partial a_t}{\partial H_j} = -(1-H_jB)^{-1}a_{t-1}$$

因此由 (7.2.18), 对于大样本情形，根的信息阵满足:

$$n^{-1}\boldsymbol{I}(\boldsymbol{G},\boldsymbol{H}) =$$

$$\left[\begin{array}{cccc:cccc} (1-G_1^2)^{-1} & (1-G_1G_2)^{-1} & \cdots & (1-G_1G_p)^{-1} & -(1-G_1H_1)^{-1} & \cdots & -(1-G_1H_q)^{-1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ (1-G_1G_p)^{-1} & (1-G_2G_p)^{-1} & \cdots & (1-G_p^2)^{-1} & -(1-G_pH_1)^{-1} & \cdots & -(1-G_pH_q)^{-1} \\ \hdashline -(1-G_1H_1)^{-1} & -(1-G_2H_1)^{-1} & \cdots & (1-G_pH_1)^{-1} & (1-H_1^2)^{-1} & \cdots & (1-H_1H_q)^{-1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ -(1-G_1H_q)^{-1} & -(1-G_2H_q)^{-1} & \cdots & -(1-G_pH_q)^{-1} & (1-H_1H_q)^{-1} & \cdots & (1-H_q^2)^{-1} \end{array}\right] \tag{7.2.21}$$

AR(2) 由此，特别对于二阶自回归过程

$$(1-G_1B)(1-G_2B)\widetilde{w}_t = a_t$$

$$\begin{aligned} \boldsymbol{V}(\hat{G}_1,\hat{G}_2) &\simeq n^{-1}\begin{bmatrix} (1-G_1^2)^{-1} & (1-G_1G_2)^{-1} \\ (1-G_1G_2)^{-1} & (1-G_2^2)^{-1} \end{bmatrix}^{-1} \\ &= \frac{1}{n}\,\frac{1-G_1G_2}{(G_1-G_2)^2}\begin{bmatrix} (1-G_1^2)(1-G_1G_2) & -(1-G_1^2)(1-G_2^2) \\ -(1-G_1^2)(1-G_2^2) & (1-G_2^2)(1-G_1G_2) \end{bmatrix} \end{aligned} \tag{7.2.22}$$

对于二阶滑动平均过程可以得到完全平行的结果。

ARMA(1,1)　类似地，对于 ARMA(1,1) 过程：

$$(1-\phi B)\widetilde{w}_t=(1-\theta B)a_t$$

在 (7.2.21) 中令 $\phi=G_1$ 和 $\theta=H_1$, 我们得到

$$\begin{aligned}\boldsymbol{V}(\hat{\phi},\hat{\theta})&=n^{-1}\begin{bmatrix}(1-\phi^2)^{-1} & -(1-\phi\theta)^{-1}\\ -(1-\phi\theta)^{-1} & (1-\theta^2)^{-1}\end{bmatrix}^{-1}\\ &=\frac{1}{n}\frac{1-\phi\theta}{(\phi-\theta)^2}\begin{bmatrix}(1-\phi^2)(1-\phi\theta) & (1-\phi^2)(1-\theta^2)\\ (1-\phi^2)(1-\theta^2) & (1-\theta^2)(1-\phi\theta)\end{bmatrix}\end{aligned}\tag{7.2.23}$$

由这两个过程得到的结果表明，自回归过程和一般的 (p,q) 阶 ARMA 过程的信息阵之间存在对偶性质. 即若与 ARMA(p,q) 过程

$$\prod_{i=1}^{p}(1-G_iB)\widetilde{w}_t=\prod_{j=1}^{q}(1-H_jB)a_t$$

相联系的参数 $(\boldsymbol{G},\boldsymbol{H})$ 的信息阵记为 $\boldsymbol{I}\{\boldsymbol{G},\boldsymbol{H}|(p,q)\}$ ，相应地，假设纯自回归模型 AR$(p+q)$

$$\prod_{i=1}^{p}(1-G_iB)\prod_{j=1}^{q}(1-H_jB)\widetilde{w}_t=a_t$$

其参数 $(\boldsymbol{G},\boldsymbol{H})$ 的信息阵记为

$$\boldsymbol{I}\{\boldsymbol{G},\boldsymbol{H}|(p+q,0)\}=\begin{bmatrix}\boldsymbol{I}_{GG} & \vdots & \boldsymbol{I}_{GH}\\ \cdots & \cdots & \cdots\\ \boldsymbol{I}'_{GH} & \vdots & \boldsymbol{I}_{HH}\end{bmatrix}$$

该矩阵在 p 行、 p 列处分块. 于是对于中等和大样本，我们可直接由 (7.2.21) 看出

$$\boldsymbol{I}\{\boldsymbol{G},\boldsymbol{H}|(p,q)\}\simeq\boldsymbol{I}\{\boldsymbol{G},-\boldsymbol{H}|(p+q,0)\}=\begin{bmatrix}\boldsymbol{I}_{GG} & \vdots & -\boldsymbol{I}_{GH}\\ \cdots & \cdots & \cdots\\ -\boldsymbol{I}'_{GH} & \vdots & \boldsymbol{I}_{HH}\end{bmatrix}\tag{7.2.24}$$

因此，对于中等和大样本情形，由于信息阵的逆给出了参数估计值协方差阵 $\boldsymbol{V}(\hat{\boldsymbol{G}},\hat{\boldsymbol{H}})$ 的一个很接近的近似，所以我们相应有

$$\boldsymbol{V}\{\hat{\boldsymbol{G}},\hat{\boldsymbol{H}}|(p,q)\} \simeq \boldsymbol{V}\{\hat{\boldsymbol{G}},-\hat{\boldsymbol{H}}|(p+q,0)\} \tag{7.2.25}$$

7.3 对具体模型的一些估计结果

在附录 A7.3 和 A7.4 中对具体情形推导出了一些估计结果。我们对于这些结果及在本章前面得到的结果加以总结，以供参考.

7.3.1 自回归过程

通过解某些线性方程组可以得到纯自回归过程的参数估计值。我们在附录 A7.4 中将要说明：

1. 怎样通过解线性方程组来得出精确的最小二乘估计 (也可见 7.4.3 节) 。

2. 怎样通过对这些方程中的系数作稍许修正，来求得精确最大似然方程的近似。

3. 怎样通过求解一系列标准线性回归模型正规方程形式的线性方程组，来求得如 7.1.3 节所定义的条件最小二乘估计。

4. 怎样用估计到的自相关系数作为 Yule–Walker 方程中的系数，以此得到与最小二乘估计和极大似然估计近似的估计值。

如同 7.2 节中用一般形式所描述的，在 (1) 中得到的估计值当然与直接最小化 $S(\phi)$ 而得出的估计值相同。估计值 (4) 是 Yule 和 Walker 所得到的著名的近似。它们作为识别阶段所使用的最初的估计是很有用的，但是在某些情形与 (1),(2) 或 (3) 的估计值可能有明显的差别。为了说明起见，我们对得自 Yule-Walker 方程的估计值 (4) 和在表 7.11 中所总结的最小二乘估计 (1) 加以比较。

Yule–Walker 估计　Yule–Walker 估计 (6.3.6) 是

$$\hat{\boldsymbol{\phi}} = \boldsymbol{R}^{-1}\boldsymbol{r}$$

其中

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{bmatrix}, \qquad \boldsymbol{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} \tag{7.3.1}$$

具体地，一阶和二阶自回归过程的估计值为

$$\begin{aligned} &\text{AR}(1): \ \hat{\phi}_1 = r_1 \\ &\text{AR}(2): \ \hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}, \qquad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2} \end{aligned} \tag{7.3.2}$$

在附录 A7.4 中说明了，$S(\hat{\boldsymbol{\phi}})$ 的一个近似可由下式提供:

$$S(\hat{\boldsymbol{\phi}}) = \sum_{t=1}^{n} \widetilde{w}_t^2 (1 - \boldsymbol{r}'\hat{\boldsymbol{\phi}}) \tag{7.3.3}$$

因此

$$\hat{\sigma}_a^2 = \frac{S(\hat{\boldsymbol{\phi}})}{n} = c_0(1 - \boldsymbol{r}'\hat{\boldsymbol{\phi}}) \tag{7.3.4}$$

这里 c_0 是 w 的样本方差。一个平行的表达式把 σ_a^2 和 γ_0 (w 的理论方差) 联系起来 (见 (3.2.8)), 即

$$\sigma_a^2 = \gamma_0(1 - \boldsymbol{\rho}'\boldsymbol{\phi})$$

这里 $\boldsymbol{\rho}$ 和 $\boldsymbol{\phi}$ 的元素都是理论值。因此由 (7.2.19), 估计值 $\hat{\boldsymbol{\phi}}$ 的协方差阵是

$$\boldsymbol{V}(\hat{\boldsymbol{\phi}}) \simeq n^{-1}\sigma_a^2 \boldsymbol{\Gamma}^{-1} = n^{-1}(1 - \boldsymbol{\rho}'\boldsymbol{\phi})\boldsymbol{P}^{-1} \tag{7.3.5}$$

如 (2.1.7) 所定义，上式中的 $\boldsymbol{\Gamma}$ 和 $\boldsymbol{P}$ 是 AR(p) 过程 p 个顺序值的自协方差和互协方差阵。

特别地，对于一阶和二阶自回归过程，我们得到

$$\text{AR}(1):V(\hat{\phi})\simeq n^{-1}(1-\phi^2) \tag{7.3.6}$$

$$\text{AR}(2):\boldsymbol{V}(\hat{\phi}_1,\hat{\phi}_2)\simeq n^{-1}\begin{bmatrix}1-\phi_2^2 & -\phi_1(1+\phi_2)\\ -\phi_1(1+\phi_2) & 1-\phi_2^2\end{bmatrix} \tag{7.3.7}$$

通过把 (7.3.5) 中的参数用估计值代替，就可以得到方差和协方差的估计。因此

$$\hat{\boldsymbol{V}}(\hat{\boldsymbol{\phi}})=n^{-1}(1-\boldsymbol{r}'\hat{\boldsymbol{\phi}})\boldsymbol{R}^{-1} \tag{7.3.8}$$

例：序列 C,D,E 和 F 已被识别为可能为自回归 (对于序列 C 是一阶差分后的自回归)，滞后为 1 或 2 的自相关列在下表中：

序列	试探识别	差分阶数	有关估计的自相关	n
C	(1,1,0)	1	r_1=0.805	225
D	(1,0,0)	0	r_1=0.861	310
E	(2,0,0)	0	r_1=0.806,r_2=0.428	100
F	(2,0,0)	0	r_1=–0.390,r_2=0.304	70

利用 (7.3.2)，我们得到在表 7.12 中列出的 Yule-Walker 估计，以及由 (7.3.6) 和 (7.3.7) 计算出的标准差。利用 (7.3.7)，二阶自回归

表 7.12 序列 C 到 F 的 Yule-Walker 估计

序列	估计值	标准差	估计之间的相关系数
C	$\hat{\phi}_1$= 0.81	±0.04	
D	$\hat{\phi}_1$= 0.86	±0.03	
E	$\hat{\phi}_1$= 1.32	±0.08	
	$\hat{\phi}_2$=–0.63	±0.08	–0.81
F	$\hat{\phi}_1$=–0.32	±0.12	
	$\hat{\phi}_2$= 0.18	±0.12	0.39

过程的两个参数估计值之间的相关系数由下式给出

$$\rho(\hat{\phi}_1,\hat{\phi}_2) \simeq \frac{\text{cov}[\hat{\phi}_1,\hat{\phi}_2]}{\sqrt{V(\hat{\phi}_1)V(\hat{\phi}_2)}} = \frac{-\phi_1}{1-\phi_2} = -\rho_1 \simeq -r_1$$

我们注意对于序列 E 的情形，估计值之间存在高度负相关，这表明 $\hat{\phi}_1$ 和 $\hat{\phi}_2$ 的置信域将沿着从西北到东南的对角线伸展。这意味着这些估计值相当不稳定，并解释了对于这一具体序列在表 7.12 中的 Yule-Walker 的估计和表 7.11 的最小二乘估计之间何以存在相对较大的差别.

7.3.2 滑动平均过程

有关滑动平均过程的极大似然估计， 在简单的情形可如节 7.1.6 中所述，通过图形得到，更一般的情形，如节 7.2.1 所述，可通过迭代计算得出。由 (7.2.20) 可知，对于适度或较大的样本，q 阶滑动平均过程参数估计值的协方差阵与同阶自回归过程相应矩阵具有相同的形式。

因此，对一阶和二阶滑动平均过程，相应于 (7.3.6) 和 (7.3.7) 有

$$\text{MA}(1):\ V(\hat{\theta}) \simeq n^{-1}(1-\theta^2) \tag{7.3.9}$$

$$\text{MA}(2):\ \boldsymbol{V}(\hat{\theta}_1,\hat{\theta}_2) \simeq n^{-1}\begin{bmatrix} 1-\theta_2^2 & -\theta_1(1+\theta_2) \\ -\theta_1(1+\theta_2) & 1-\theta_2^2 \end{bmatrix} \tag{7.3.10}$$

7.3.3 混合过程

与滑动平均过程一样，混合过程的极大似然估计 $(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})$ 在简单情形可以由图形得出，对更一般的情形可通过递推计算求得。对于中等及大样本情形，协方差阵可由信息阵 (7.2.18) 的计算与求逆得到。对一些重要的特殊情形，如 ARMA(1,1) 过程

$$(1-\phi B)\widetilde{w}_t = (1-\theta B)a_t$$

如同 (7.2.23), 我们得到

$$V(\hat{\phi},\hat{\theta}) \simeq n^{-1}\frac{1-\phi\theta}{(\phi-\theta)^2}\begin{bmatrix}(1-\phi)^2(1-\phi\theta) & (1-\phi^2)(1-\theta^2)\\(1-\phi^2)(1-\theta^2) & (1-\theta^2)(1-\phi\theta)\end{bmatrix} \tag{7.3.11}$$

我们注意到，当 $\phi=\theta$ 时，$\hat{\phi}$ 和 $\hat{\theta}$ 的方差是无穷大. 这是可以料想的，因为在这种情形下，因子 $(1-\phi B)=(1-\theta B)$ 在模型两边可以消去，于是模型成为

$$\widetilde{w}_t = a_t$$

这是参数冗余的特例，我们将在 7.3.5 节中更充分地讨论.

7.3.4 在估计中线性和非线性分量的分离*

有时进行这样的分析是有意义的：将混和模型参数的估计分成基本线性和非线性两个部分. 考虑一般混合模型，我们写作

$$a_t = \phi(B)\theta^{-1}(B)\widetilde{w}_t$$

或

$$a_t = \phi(B)(\epsilon_t|\boldsymbol{\theta}) \tag{7.3.12}$$

其中

$$(\epsilon_t|\boldsymbol{\theta}) = \theta^{-1}(B)\widetilde{w}_t$$

即

$$\widetilde{w}_t = \theta(B)(\epsilon_t|\boldsymbol{\theta}) \tag{7.3.13}$$

对于任意给定的一组 θ,ϵ 可以有 (7.3.13) 递推计算，可记为

$$\epsilon_t = \widetilde{w}_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$$

该递推可以由设未知的 ϵ 为 0 而开始. 在算出 ϵ 后，条件估计 $\hat{\phi}_0$ 便可立刻得到. 这些都是线性模型 (7.3.12) 中待估计的自回归参数，该模型可记为

$$a_t = \epsilon_t - \phi_1\epsilon_{t-1} - \phi_2\epsilon_{t-2} - \cdots - \phi_p\epsilon_{t-p} \tag{7.3.14}$$

---

* 从此至 7.3.5 前的内容在初次阅读时可以略去.

正如我们在 7.3.1 节中所解释的，自回归参数的最小二乘估计可以通过直接求解简单的线性方程组而得到，并可以用 Yule-Walker 方程来近似。在简单的情形，我们可以在 $\boldsymbol{\theta}$ 的网格点上计算 $S(\hat{\boldsymbol{\phi}}_0,\boldsymbol{\theta})$ 并绘出等值线图，以此考察 $S(\hat{\boldsymbol{\phi}}_0,\boldsymbol{\theta})$ 的性质并求得绝对的最小平方和。

用于序列 C 的例子　根据在表 6.4 中给出对序列 C 所作试探性识别，可知该序列可能是由 (1,1,0) 模型

$$(1-\phi B)w_t = a_t$$

生成，其中 $w_t=\nabla z_t, E[w_t]=0$ ．确定该模型是为了考察数据对于更复杂模型

$$(1-\phi B)w_t=(1-\theta_1 B-\theta_2 B^2)a_t$$

的估计态势。由上述讨论得知，可以认为该过程是由非线性模型

$$\epsilon_t=w_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}$$

和线性模型

$$a_t=\epsilon_t-\phi\epsilon_{t-1}$$

组合生成的。对于可逆域内非线性参数 $\boldsymbol{\theta}=(\theta_1,\theta_2)$ 的每种选择，递推算出一组 ϵ，然后由 Yule-Walker 近似可以得到估计值 $\hat{\phi}_0=r_1(\epsilon)$，并同时得出

$$S(\hat{\phi}_0,\boldsymbol{\theta})=\sum_{t=1}^{n}\epsilon_t^2[1-r_1^2(\epsilon)]$$

将该平方和在 θ_1 和 θ_2 的网格点上标出，它的等值线如图 7.6 中所示。我们看到，最小值接近于图上标出的 $\theta_1=\theta_2=0$ 的点，在该点上 $r_1(\epsilon)=0.850$ ．因此，可以确认，在 (1,1,2) 阶的全体模型类中，简单的 (1,1,0) 模型 $(1-0.8B)\nabla z_t=a_t$ 提供了最恰当的表示。

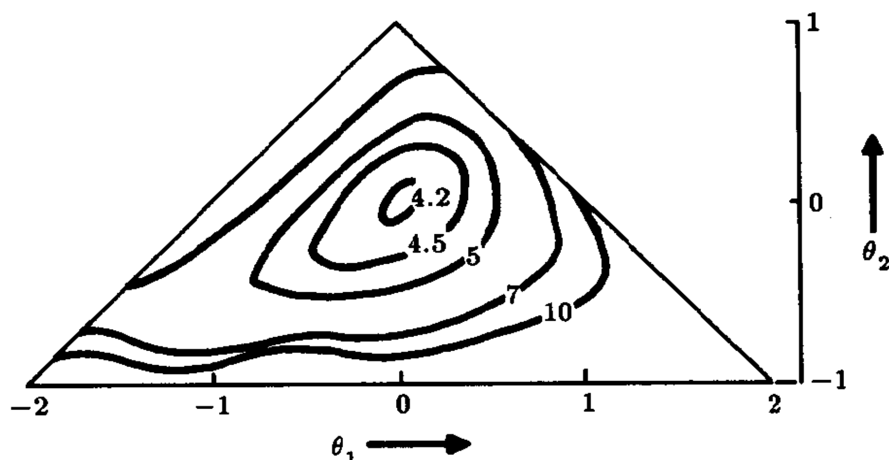


图 7.6 在参数 θ 容许空间上画出的 $S(\hat{\phi}_0,\boldsymbol{\theta})$ 等值线图

7.3.5 参数冗余

模型

$$\phi(B)\widetilde{w}_t = \theta(B)a_t$$

等同于模型

$$(1-\alpha B)\phi(B)\widetilde{w}_t = (1-\alpha B)\theta(B)a_t$$

这里，自回归和滑动平均算子都乘以相同的因子 $1-\alpha B$. 如果在模型拟合时包含了冗余的子，则在估计过程中会出现严重的困难. 因此，需要多加小心，避免出现冗余或接近冗余因子的情形. 还需要注意到，冗余因子的存在不一定是明显的。例如，我们来看一下 ARMA(2,1) 模型

$$(1-1.3B+0.4B^2)\widetilde{w}_t = (1-0.5B)a_t$$

的公共因子，只有对左边进行分解因式才得出

$$(1-0.5B)(1-0.8B)\widetilde{w}_t = (1-0.5B)a_t$$

从而有

$$(1-0.8B)\widetilde{w}_t = a_t$$

在实际中，不仅是精确相等的因子会造成麻烦，近似相等的因子也是如此. 例如，假设真实模型是

$$(1-0.4B)(1-0.8B)\widetilde{w}_t = (1-0.5B)a_t \tag{7.3.15}$$

如果我们将模型拟合为 ARMA(2,1), 则由于模型左、右两边存在近似相等的因子 $(1-0.4B)$ 和 $(1-0.5B)$, 参数估计的不稳定性是可以想见的。在这种情形我们会发现，不同参数值的组合将得出相近的 $[a]$ 和相近的似然函数，而且左边参数值的变化可由右边参数的适当变化来补偿。在三维参数空间中，平方和等值曲面接近于倾斜走向的柱面，我们能够找到的不是清晰定义的最小值点，而是成直线态势的“近似最小二乘”。

从一个稍有不同的观点来看，我们可以把模型 (7.3.15) 用无穷阶的自回归算子表示。作必要的展开后，我们发现：

$$(1-0.700B-0.030B^2-0.015B^3-0.008B^4-\cdots)\widetilde{w}_t=a_t$$

因此，该模型非常接近于

$$(1-0.7B)\widetilde{w}_t=a_t \tag{7.3.16}$$

当我们试图用 ARMA(2,1) 模型去拟合时，所得的估计值会出现不稳定性，这是因为当用一个参数就可以表示的情形下，我们却尝试用三个参数去拟合。

预先识别 — 避免参数冗余的一种方法　在模型拟合之前要进行模型识别，这样做的主要原因是为了避免因参数冗余而引起的困难。或者说是为了更准确可靠，为了实现参数化的简约性。因此在刚才考虑的例子中，对仅有几百个观测值的时间序列，可以估计出其自相关函数，但相应的数据是由 (7.3.15) 产生，还是由简单的自回归过程 (7.3.16) 产生，实在是难以区分。

所有这些导致我们去拟合 AR(1) 过程，通常这样做就足够了。只有当时间序列有数千观侧值时才需要去检测更复杂的模型。在这种情形下，才能有充足的信息来得到额外参数的较好估计。

ARMA(1,1) 过程中的冗余　有可能发生因子直接相约的最简单的过程是 ARMA(1,1) 过程：

$$(1-\phi B)\widetilde{w}_t=(1-\theta B)a_t$$

特别地，如果 $\phi=\theta$，无论它们有什么样的值，上式都成为

$$\widetilde{w}_t = a_t$$

该模型蕴含 w_t 是由一个白噪声过程产生。因此，数据不能提供一般参数的信息，由 (7.3.11)，$\hat{\phi}$ 和 $\hat{\theta}$ 有无穷大的方差。而且，无论 ϕ 和 θ 为何值，$S(\phi,\theta)$ 在直线 $\phi=\theta$ 上必然为常数，就象图 7.7 中的直线，该图是对序列 A 作出的平方和函数图形。数据的最小二乘估计 $\hat{\phi}=0.92$ 和 $\hat{\theta}=0.58$ 所对应的点并不特别接近于直线 $\phi=\theta$，在估计这些参数时不会发生困难。

在实际中，如果采用了我们推荐的识别技巧，这些困难可以避免。若在 ARMA(1,1) 过程中 ϕ 接近于 θ，则通常将被识别为白噪声，或者，若两个参数之间的差别是可感知的，则将被识别为具有一个很小系数的 AR(1) 或 MA(1) 过程。总之：

1. 我们应该避免包含近似公因子的混合过程，且应提防由此而造成的困难。

2. 如果我们正确地使用了识别和估计方法，就能自动地避免了冗余参数的处置。

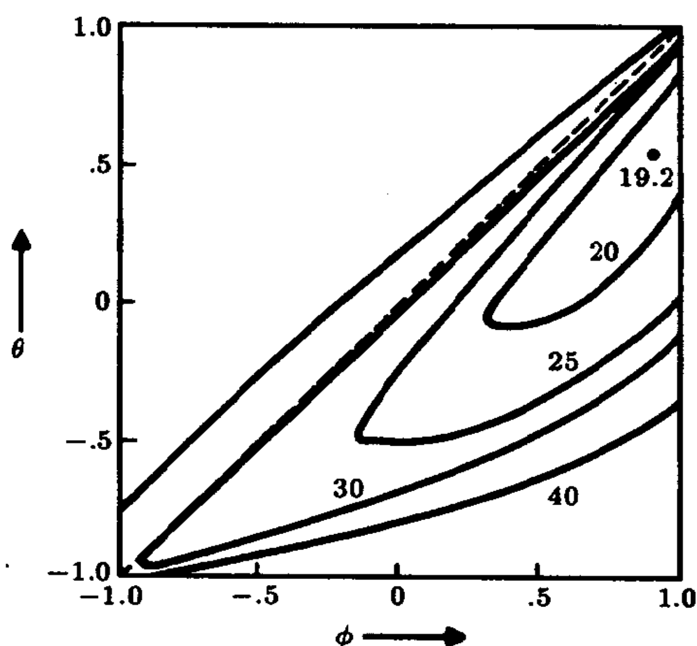


图 7.7 序列 A 的平方和图形

7.4 使用 Bayes 原理的估计 *

7.4.1 Bayes 原理

在本节中，我们仍用 ξ 表示一个一般的参数向量。Bayes 原理告诉我们，若 $p(\xi)$ 是数据采集之前 ξ 的先验概率分布，$p(\xi|z)$ 是 ξ 关于数据 z 的后验分布，$p(\xi|z)$ 可通过把先验分布 $p(\xi)$ 和似然函数 $L(\xi|z)$ 联合起来而得到，具体的形式为

$$p(\xi|z) = \frac{p(\xi)L(\xi|z)}{\int p(\xi)L(\xi|z)d\xi} \tag{7.4.1}$$

分母只是为了保证 $p(\xi|z)$ 的积分为 1. 这个表达式的重要部分是分子，由此我们可以看到，后验分布与先验分布和似然函数的乘积成比例。Savage[173] 已说明了先验和后验概率可被解释为主观概率。特别地，通常在数据可利用之前我们对 ξ 掌握的信息非常少，我们只能认为在相关的范围内 ξ 所显示出的先验信息是：既可能是某一值，又可能是另一值。在这种情形，$p(\xi)$ 可取为局部均匀分布，因此 $p(\xi|z)$ 与似然函数成正比。

应该注意到，为了使该结论成立，ξ 的先验分布不一定在其整个范围内 (对于某些参数这一范围可以是无限的) 是均匀分布。局部均匀的要求是指：在似然函数可被估计的范围内近似是均匀的，而在范围之外不会取特别大的值。

因此，如果 ξ 是一把椅子的重量，我们当然可以明确地说出这样的先验信息：它的重量超过 1 盎司，但少于 1 吨。还可能是这样：当我们在一个称量器上称这把椅子时，得到一个观侧值 z 及误差的标准差 σ, 我们就可以老老实实地说：对于在 $z \pm 3\sigma$ 范围内的先验值我们都会同样地满意。例外的情形是，如果称量器称出一把明显很重的椅子，比如重量为 10 盎司。在这种情形，似

---

*7.4 节的内容在初次阅读时可以略去.

然函数与先验性质是大相径庭的。因此我们当然不会去用 Bayes 原理来组合它们，而是检查称重器，若它是准确的，则应更仔细地检查椅子。

当然在上述思想中存在某种武断性。假定我们设 ξ 的先验分布是局部均匀的，于是这蕴含 ξ 的任意线性函数的分布也是局部均匀的。然而一些非线性变换 $\alpha = \alpha(\xi)$，例如 $\alpha = \log\xi$，其先验分布可能不是精确局部均匀的。如果我们能得到 ξ 相当准确的估计，则这种武断性通常几乎没有影响。那么，我们仅在一个小区域上考虑 ξ，在该区域上对于 ξ 的变换，例如 $\log\xi$，通常将会非常接近于线性的。

Jeffreys[117] 曾指出：最好选择度量 $\alpha(\xi)$ 以使 Fisher 信息测度 $I_\alpha = -E[\partial^2 l/\partial\alpha^2]$ 与值 α 及 ξ 独立。这等价于选择 $\alpha(\xi)$ 以使其极大似然估计的极限方差与 ξ 独立，这些是通过选择与 $\sqrt{I_\xi}$ 成比例的先验分布来得到的。

Jeffreys 在对所使用的参数化方法不变的基础上证明了这一先验选择。特别地，对于这种选择， $\alpha(\xi)$ 和 ξ 的后验分布满足 $p(\xi|\boldsymbol{z}) = p(\alpha|\boldsymbol{z})d\alpha/d\xi$，这里， $\alpha(\xi)$ 和 ξ 由 $1:1$ 的变换相联系。由下述讨论可得同样的结果 [58] 。如果对于大样本， $\alpha(\xi)$ 的期望似然函数接近于正态曲线，于是该曲线的期望与方差概括了可望由数据得到的信息。现在假设可以找到一个变换 $\alpha(\xi)$，对于它来说，不管参数的真值是什么，用来近似的正态曲线都有接近于常数的方差。那么，在参数化中，可望来自于数据的唯一信息是由预期似然函数的位置来传达的。所谓我们对这个预期观测信息的有关先验一无所知，就是说我们认为 α 的不同位置有同样可能的先验信息。等价地，就是我们所谓的 α 应取为局部均匀的。

Jeffrey 准则对处理多个参数的推广是用参数 ξ 的联合先验分布，将其取为与下式成比例：

$$|\boldsymbol{I}_\xi|^{1/2} = \left|-E\left[\frac{\partial l}{\partial\xi_i\partial\xi_j}\right]\right|^{1/2} \tag{7.4.2}$$

在 [119] 中指出，采用这样的度量 α，使得似然函数得到了最好的考虑与设计，这时 I_α 独立于 α。有了以上说明，我们注意到，在均匀先验的假设下，似然函数和后验密度函数是成比例的。

7.4.2 参数的 Bayes 估计

现在我们从 Bayes 观点来考虑 ARIMA 模型的参数估计。附录 A7.3 证明了：来自 ARIMA(p,d,q) 过程长度为 $N=n+d$ 的时间序列 $\boldsymbol{z}$ 其精确似然函数的形式为

$$L(\boldsymbol{\phi},\boldsymbol{\theta}|\boldsymbol{z})=\sigma_a^{-n}f(\boldsymbol{\phi},\boldsymbol{\theta})\exp\left[-\frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{2\sigma_a^2}\right] \tag{7.4.3}$$

其中

$$S(\boldsymbol{\phi},\boldsymbol{\theta})=\sum_{t=1}^{n}[a_t|\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}]^2+[\boldsymbol{e}_*]'\boldsymbol{\Omega}^{-1}[\boldsymbol{e}_*] \tag{7.4.4}$$

如果我们没有关于 $\sigma_a^2,\boldsymbol{\phi}$ 或 $\boldsymbol{\theta}$ 的先验信息，并且因为 σ_a^2 的信息不会给出 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的信息，则如 Jeffrey 所述，使用如下形式的 $\boldsymbol{\phi},\boldsymbol{\theta}$ 和 σ_a^2 的先验分布是合理的：

$$p(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a^2)\propto|\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})|^{1/2}\sigma_a^{-1}$$

于是可得后验分布为

$$p(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a^2|\boldsymbol{z})\propto\sigma_a^{-(n+1)}|\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})|^{1/2}f(\boldsymbol{\phi},\boldsymbol{\theta})\exp\left[-\frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{2\sigma_a^2}\right] \tag{7.4.5}$$

如果我们现在对 (7.4.5) 关于 σ_a 从 0 到无穷积分，我们就可得到参数 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 精确的联合后验分布：

$$p(\boldsymbol{\phi},\boldsymbol{\theta}|\boldsymbol{z})\propto|\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})|^{1/2}f(\boldsymbol{\phi},\boldsymbol{\theta})\{S(\boldsymbol{\phi},\boldsymbol{\theta})\}^{-n/2} \tag{7.4.6}$$

7.4.3 自回归过程

若 z_t 适合于一个 $(p,d,0)$ 过程，则 $w_t=\nabla^d z_t$ 适合于 p 阶纯自回归过程。在附录 A7.4 中指出，在任何情形下，$|\boldsymbol{I}(\boldsymbol{\phi})|^{1/2}$ 和 $f(\boldsymbol{\phi})$ 都被 $S(\boldsymbol{\phi})$ 中的项所控制，实质上被消去了。这样就得出了非常简单的结果：在给定假设之下，w 的 AR(p) 过程其参数 $\boldsymbol{\phi}$ 具有后验分布为

$$p(\boldsymbol{\phi}|\boldsymbol{z}) \propto \{S(\boldsymbol{\phi})\}^{-n/2} \tag{7.4.7}$$

由此结论可知，当先验信息一无所知时，平方和的等值线 (它近似于似然函数等值线) 就是后验概率的等值线。

自回归参数的联合分布　在附录 A7.4 中已证明：对于纯自回归过程，使 $S(\boldsymbol{\phi})=\boldsymbol{\phi}_u'\boldsymbol{D}\boldsymbol{\phi}_u$ 极小化的 $\boldsymbol{\phi}$ 的最小二乘估计为

$$\hat{\boldsymbol{\phi}}=\boldsymbol{D}_p^{-1}\boldsymbol{d} \tag{7.4.8}$$

这里，

$$\boldsymbol{\phi}_u=\begin{bmatrix}1\\ \cdots\\ \boldsymbol{\phi}\end{bmatrix}\quad \boldsymbol{d}=\begin{bmatrix}D_{12}\\ D_{13}\\ \vdots\\ D_{1,p+1}\end{bmatrix}\quad \boldsymbol{D}_p=\begin{bmatrix}D_{22} & D_{23} & \cdots & D_{2,p+1}\\ D_{23} & D_{33} & \cdots & D_{3,p+1}\\ \vdots & \vdots & \cdots & \vdots\\ D_{2,p+1} & D_{3,p+1} & \cdots & D_{p+1,p+1}\end{bmatrix}$$

$$\boldsymbol{D}=\begin{bmatrix}D_{11} & \vdots & -\boldsymbol{d}'\\ \cdots & \cdots & \cdots\\ -\boldsymbol{d} & \vdots & \boldsymbol{D}_p\end{bmatrix} \tag{7.4.9}$$

及

$$D_{ij}=D_{ji}=\tilde{w}_i\tilde{w}_j+\tilde{w}_{i+1}\tilde{w}_{j+1}+\cdots+\tilde{w}_{n+1-i}\tilde{w}_{n+1-j} \tag{7.4.10}$$

于是有

$$S(\boldsymbol{\phi})=\nu s_a^2+(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}})'\boldsymbol{D}_p(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}) \tag{7.4.11}$$

这里

$$s_a^2 = \frac{S(\hat{\boldsymbol{\phi}})}{\nu} \quad , \quad \nu = n - p \tag{7.4.12}$$

及

$$S(\boldsymbol{\phi}) = \hat{\boldsymbol{\phi}}_u' \boldsymbol{D} \hat{\boldsymbol{\phi}}_u = D_{11} - \hat{\boldsymbol{\phi}}' \boldsymbol{D}_p \hat{\boldsymbol{\phi}} = D_{11} - \boldsymbol{d}' \boldsymbol{D}_p^{-1} \boldsymbol{d} \tag{7.4.13}$$

因此我们可以写成

$$p(\boldsymbol{\phi}|\boldsymbol{z}) \propto \left[1 + \frac{(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})' \boldsymbol{D}_p (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})}{\nu s_a^2} \right]^{-n/2} \tag{7.4.14}$$

等价地有

$$p(\boldsymbol{\phi}|\boldsymbol{z}) \propto \left[+ \frac{\frac{1}{2} \sum_i \sum_j S_{ij} (\phi_i - \hat{\phi}_i)(\phi_j - \hat{\phi}_j)}{\nu s_a^2} \right]^{-n/2} \tag{7.4.15}$$

其中

$$S_{ij} = \frac{\partial^2 S(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j} = 2 D_{i+1, j+1}$$

于是得到自回归参数的后验概率, 它是 (A7.1.13) 的多维 t- 分布, 自由度为 $\nu = n - p$.

具体地, 对 $p = 1$ 的特例, $(\phi - \hat{\phi})/s_{\hat{\phi}}$ 恰好是自由度为 $n - 1$ 的学生 t- 分布, 利用上面出的一般结果, $\hat{\phi}$ 和 $s_{\hat{\phi}}$ 给出为

$$\hat{\phi} = \frac{D_{12}}{D_{22}} \qquad s_{\hat{\phi}} = \left[\frac{1}{(n-1)} \frac{D_{11}}{D_{22}} \left(1 - \frac{D_{12}^2}{D_{11} D_{22}} \right) \right]^{1/2} \tag{7.4.16}$$

对于大样本情形, 量 $s_{\hat{\phi}}$ 趋于 $[(1 - \phi^2)/n]^{1/2}$, 且在抽样理论框架中等同于 $\hat{\phi}$ 的大样本 "标准差" 。但是在 Bayes 框架中使用该式及类似的表达式时, 应该记住参数 (此例中是 ϕ) 是随机变量。象 $\hat{\phi}$ 和 $s_{\hat{\phi}}$ 这样的量是已发生数据的函数, 它们是固定不变的。

正态近似　对于样本量 $n > 50$— 这通常是我们感兴趋的情形，用正态分布类近似 t 分布是充分的。因此，$\boldsymbol{\phi}$ 非常接近于 p 变量的联合正态分布 $N(\hat{\boldsymbol{\phi}}, \boldsymbol{D}_p^{-1}s_a^2)$，其均值为 $\hat{\boldsymbol{\phi}}$，方差和协方差阵为 $\boldsymbol{D}_p^{-1}s_a^2$。

最高概率密度的 Bayes 区域　让我们来总结一下，有关各种 $\boldsymbol{\phi}$ 的值后验分布到底告诉我们什么信息，为此，指出所谓的高概率密度区域是很有用的，它也简称 [61] 为 HPD 区域。一个 Bayes 的 $1-\epsilon$ 区域有如下性质：

1. 任何区域内的参数点比区域外的点有较高的概率密度。
2. 区域内后验概率总量为 $1-\epsilon$。

由于 $\boldsymbol{\phi}$ 具有多维 t- 分布，所以由 (A7.1.4) 的结果得到

$$\Pr\{(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}})'\boldsymbol{D}_p(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}) < ps_a^2F_\epsilon(p,\nu)\} = 1-\epsilon \tag{7.4.17}$$

它定义了 $\boldsymbol{\phi}$ 精确的 HPD 区域。现在，对于 $\nu = n-p > 100$,

$$pF_\epsilon(p,\nu) \simeq \chi_\epsilon^2(p)$$

另外，

$$(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}})'\boldsymbol{D}_p(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}) = \frac{1}{2}\sum_{i,j} S_{ij}(\phi_i-\hat{\phi}_i)(\phi_j-\hat{\phi}_j)$$

因此，(7.4.17) 定义的 HPD 区域使得近似有

$$\sum_{i,j}\boldsymbol{S}_{ij}(\phi_i-\hat{\phi}_i)(\phi_j-\hat{\phi}_j) < 2s_a^2\chi_\epsilon^2(p) \tag{7.4.18}$$

若我们令 $\hat{\sigma}_a^2 = s_a^2$，它等同于 (7.1.26) 所定义的置信域。

尽管这些近似区域是相同的，但应记住它们的解释是不同的。从抽样理论的观点来看，，如果一个置信域是根据 (7.1.26) 计算的，那么对于每一组重复样本而言，这些区域将以比例 $1-\epsilon$

包含真正的参数点。从 Bayes 观点来看，我们所关心的只是具体观测到的单一样本 $\boldsymbol{z}$ 。假设我们所取的是无信息的先验分布，那么，当给定 $\boldsymbol{z}$ 后， HPD 区域中相应的 $\boldsymbol{\phi}$ 所生成的概率分布中密度最大的占比例为 $1-\epsilon$ 。换言之，使数据 $\boldsymbol{z}$ 发生的 $\boldsymbol{\phi}$ 值位于 HPD 区域内的概率为 $1-\epsilon$ 。

利用 (7.4.11),(7.4.12) 和 (7.4.18)，对于大样本时， Bayes HPD 区域是以下面曲面的等值线作为边界：

$$S(\boldsymbol{\phi}) = S(\hat{\boldsymbol{\phi}})\left[1+\frac{\chi_\epsilon^2(p)}{n}\right] \tag{7.4.19}$$

它完全对应于 (7.1.28) 定义的置信区域。

7.4.4 滑动平均过程

若 z 适合于一个 $(0,d,q)$ 阶求和滑动平均过程，那么， $w=\nabla^d z$ 便适合 q 阶纯滑动平均过程。由于自回归模型和滑动平均模型之间的估计结果，特别是信息阵存在对偶性，因此对于滑动平均情形得到：在 (7.4.6) 中的因子 $|\boldsymbol{I}(\boldsymbol{\theta})|^{1/2}$ 和 $f(\boldsymbol{\theta})$ 在任何情况下都被 $S(\boldsymbol{\theta})$ 所控制，当大样本时也可以消去。对应于 (7.4.7)，我们求得 w 的 MA(q) 过程其参数 $\boldsymbol{\theta}$ 的后验分布为

$$p(\boldsymbol{\theta}|\boldsymbol{z}) \propto [S(\boldsymbol{\theta})]^{-n/2} \tag{7.4.20}$$

对于适度样本，平方和等值线仍然基本上为后验密度的精确等值线。然而，由于 $[a_t]$ 不是 $\boldsymbol{\theta}$ 的线性函数，故 $S(\boldsymbol{\theta})$ 不会是 $\boldsymbol{\theta}$ 精确的二次型。尽管对大样本情形，在相关的区域内通常是接近二次的。这时，我们近似有

$$S(\boldsymbol{\theta}) = \nu\sigma_a^2 + \frac{1}{2}\sum_{i,j} S_{ij}(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)$$

其中 $\nu s_a^2 = S(\hat{\boldsymbol{\theta}}), \nu = n-q$ 。在 (7.4.20) 中替代 $S(\boldsymbol{\theta})$ 并利用指数近似，我们得到：

1. 对于大样本，$\boldsymbol{\theta}$ 近似为多维正态分布 $N\{\hat{\boldsymbol{\theta}}, 2\{\boldsymbol{S}_{ij}\}^{-1}s_a^2\}$。

2. 用 q 代替 p,$\boldsymbol{\theta}$ 代替 $\boldsymbol{\phi}$, 由 (7.4.18) 或 (7.4.19) 定义了近似的 HPD 区域。

例：对于 IMA(0,1,1) 过程参数 $\boldsymbol{\lambda}=1-\theta$ 的后验分布　为了说明起见，表 7.13 列出了由序列 B 得到的后验分布密度 $p(\lambda|z)$ 的计算．表中的第二列给出了在长度为 $h=0.025$ 的 λ 的间隔上计算出的纵坐标 $[S(\lambda)]^{-n/2}\times 10^{57}$．它们的和 Σ 是 236,325. 用 $h\Sigma$ 去除坐标值，我们得出后验概率密度函数。在充分近似下，该

表 7.13　对序列 B 的近似后验密度 $p(\lambda|z)$ 的计算

λ	$10^{57}\times[S(\lambda)]^{-184.5}$	$p(\lambda\|z)$
1.300	4	0.001
1.275	33	0.006
1.250	212	0.036
1.225	1,007	0.171
1.200	3,597	0.609
1.175	9,956	1.685
1.150	21,195	3.582
1.125	34,762	5.884
1.100	44,473	7.528
1.075	44,988	7.615
1.050	35,835	6.066
1.025	22,563	3.819
1.000	11,277	1.908
0.975	4,540	0.769
0.950	1,457	0.247
0.925	372	0.062
0.900	76	0.012
0.875	12	0.002
0.850	2	0.000
	236,325	40.000

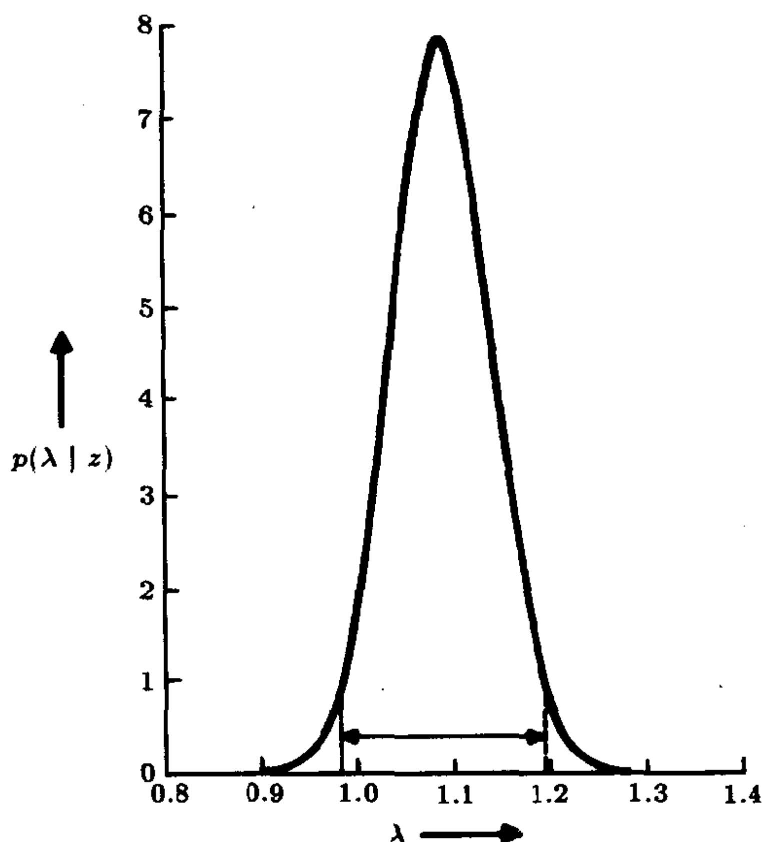


图 7.8 序列 B 的后验密度 $p(\lambda|z)$

曲线下的面积等于 1[①]．

在图 7.8 中画出了这一分布．可看出它是近似正态的，众数为 $\hat{\lambda}=1.09$, 标准差约为 0.05． 95% 的 Bayes 区间近似为 $0.98<\lambda<1.19$, 置信区间差不多是同一区间．

7.4.5 混合过程

若 z 适合于一个 (p,d,q) 阶的 ARIMA 过程，那么， $w=\nabla^d z$ 适合 ARMA(p,q) 过程 $\phi(B)\widetilde{w}_t=\theta(B)a_t$．可以证明，对这样一个过程 (7.4.5) 中的因子 $|\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})|^{1/2}$ 和 $f(\boldsymbol{\phi},\boldsymbol{\theta})$ 不会完全抵销．相反，根据 (7.4.5) 我们可以证明

$$|\boldsymbol{I}(\boldsymbol{\phi},\boldsymbol{\theta})|^{1/2}f(\boldsymbol{\phi},\boldsymbol{\theta})=J(\boldsymbol{\phi}^*|\boldsymbol{\phi},\boldsymbol{\theta}) \tag{7.4.21}$$

① 用近似的数值积分来获得一条单位面积的"标准化"曲线，如果需要的话，可以更精确地进行，使得可以读出具体的概率密度．对于几乎所有的实际应用，这一修正并不需要．计算并在某个方便的标度上画出 $[S(\lambda)]^{-n/2}$ 就足够了，这与实际的密度成比例．

在 (7.4.21) 中，$\boldsymbol{\phi}^*$ 是自回归和滑动平均算子相乘所得的 $p+q$ 个参数，有

$$(1-\phi_1^*B-\phi_2^*B^2-\cdots-\phi_{p+q}^*B^{p+q})$$
$$=(1-\phi_1B-\cdots-\phi_pB^p)\times(1-\theta_1B-\cdots-\theta_qB^q)$$

且 $\boldsymbol{J}$ 是从 $\boldsymbol{\phi}^*$ 到 $(\boldsymbol{\phi},\boldsymbol{\theta})$ 变换的 Jacobi 行列式，即

$$p(\boldsymbol{\phi},\boldsymbol{\theta}|\boldsymbol{z})\propto \boldsymbol{J}(\boldsymbol{\phi}^*|\boldsymbol{\phi},\boldsymbol{\theta})[S(\boldsymbol{\phi},\boldsymbol{\theta})]^{-n/2} \qquad (7.4.22)$$

特别对于 ARMA(1,1) 过程，$\phi_1^*=\phi+\theta, \phi_2^*=-\phi\theta, J=|\phi-\theta|$，且

$$p(\phi,\theta|\boldsymbol{z})\propto|\phi-\theta|[S(\phi,\theta)]^{n/2} \qquad (7.4.23)$$

这时，可看到在靠近直线 $\phi=\theta$ 的区域内， Jacobi 行列式将起主导作用，并在该直线上得到零密度。这是合理的，因为平方和 $S(\phi,\theta)$ 对于任何 $\phi=\theta$ 取有限值 $\sum\limits_{t=1}^{2}\widetilde{w}_t^2$，这蕴含 $\widetilde{w}_t$ 是白噪声的可能性。然而在推导中，我们并没有限制参数的区域。因此，$\phi=\theta$ 的可能性与 (相等) 参数的无限制区域相联系。限制参数空间，例如引入平稳性和可逆性 $(-1<\phi<1,-1<\theta<1)$，其结果是使密度得到一个较小的正值，但是这一修改看来几乎是没什么价值的。

Bayes 分析更加强了在 7.3.5 节中给出的观点：对混合模型进行参数估计时将会遇到困难，特别是当参数近乎冗余而又需要迭代求解时，更是如此。正如我们已经知道的，使用初步识别常常可以确保避免这种情形发生。

7.5 基于状态空间模型的似然函数

我们来回忆一下在 5.5 节中已经指出的： ARMA 过程的状态空间表示和 Kalman 滤波方法对于精确的有限样本预测和精确的似然函数分析是很有用的。这种方法已由 Jones[123], Gardner, Harvey 和 Phillips [94] 及其它一些人描述过。现在我们来说明：如何利用状态空间模式中的 Kalman 滤波作为计算精确似然函数

的方便方法. 我们再次设定, 来自 ARMA(p,q) 模型的 n 个给定观测值为 $w_1, w_2, \cdots, w_n$ 。

如 5.5 节所述, 对于 ARMA(p,q) 模型 $\phi(B)w_t = \theta(B)a_t$, 状态空间模型形式由 $\boldsymbol{Y}_t = \boldsymbol{\Phi}\boldsymbol{Y}_{t-1} + \boldsymbol{\Psi}$ 和 $w_t = \boldsymbol{H}\boldsymbol{Y}_t$ 给出, 其中 $\boldsymbol{Y}'_t = (w_t, \hat{w}_t(1), \cdots, \hat{w}_t(r-1)), r = max(p, q+1)$,$\boldsymbol{H} = (1, 0, \cdots, 0)$,

$$\boldsymbol{\Phi} = \begin{bmatrix} 0, & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 1 \\ \phi_r & \phi_{r-1} & \cdots & \cdots & \phi_1 \end{bmatrix}$$

及 $\boldsymbol{\Psi} = (1, \psi_1, \cdots, \psi_{r-1})$ 。 Kalman 滤波方程 (5.5.5)-(5.5.8) 给出了状态向量 $\boldsymbol{Y}_t$ 的一步外推预报值为 $\hat{\boldsymbol{Y}}_{t|t-1} = E[\boldsymbol{Y}_t | w_{t-1}, \cdots, w_1]$, 以及误差协方差阵 $\boldsymbol{V}_{t|t-1} = E[(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1})(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1})']$ 。特别地, 对于 ARMA(p,q) 模型的状态空间形式, 对于 $t = 1, 2, \cdots, n$, 递推方程为

$$\hat{\boldsymbol{Y}}_{t|t} = \hat{\boldsymbol{Y}}_{t|t-1} + \boldsymbol{K}_t(w_t - \hat{w}_{t|t-1}) \quad (7.5.1)$$

其中 $\boldsymbol{K}_t = \boldsymbol{V}_{t|t-1}\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{V}_{t|t-1}\boldsymbol{V}']^{-1}$, 及 $\hat{w}_{t|t-1} = \boldsymbol{H}\hat{\boldsymbol{Y}}_{t|t-1}$, 且

$$\hat{\boldsymbol{Y}}_{t|t-1} = \boldsymbol{\Phi}\hat{\boldsymbol{Y}}_{t-1|t-1} \qquad \boldsymbol{V}_{t|t-1} = \boldsymbol{\Phi}\boldsymbol{V}_{t-1|t-1}\boldsymbol{\Phi}' + \sigma_a^2\boldsymbol{\Psi}\boldsymbol{\Psi}' \quad (7.5.2)$$

其中

$$\boldsymbol{V}_{t|t} = [\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H}]\boldsymbol{V}_{t|t-1} \quad (7.5.3)$$

以上对于 $= t = 1, 2, \cdots, n$ 成立。特别 , 预报向量的第一分量是 $\hat{w}_{t|t-1} = \boldsymbol{H}\hat{\boldsymbol{Y}}_{t|t-1} = E[w_t | w_{t-1}, \cdots, w_1]$, 元素 $\sigma_a^2 v_t = \boldsymbol{H}\boldsymbol{V}_{t|t-1}\boldsymbol{H}' = E[(w_t - \hat{w}_{t|t-1})^2]$ 是预报误差方差。

为了用上面的结果得到 n 个观测值的向量 $\boldsymbol{w}' = (w_1, w_2, \cdots, w_n)$ 的精确似然函数, 我们注意到 w 的联合分布被因子分解为

$$p(\boldsymbol{w}|\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_a) = \prod_{t=1}^{n} p(w_1 | w_{t-1}, \cdots, w_1, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_a) \quad (7.5.4)$$

其中 $p(w_t|w_{t-1},\cdots,w_1,\boldsymbol{\phi},\boldsymbol{\theta},\sigma_1)$ 表示给定 $w_{t-1},\cdots,w_1$ 时 w_t 的条件分布. 在正态假定下, 该条件分布是具有 (条件) 均值 $\hat{w}_{t|t-1}=E[w_t|w_{t-1},\cdots,w_1]$ 和 (条件) 方差 $\sigma_a^2 v_t = E[(w_t-\hat{w}_{t|t-1})^2]$ 的正态. 因此, $\boldsymbol{w}$ 的联合分布可以方便地表示为

$$p(\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a)=\prod_{t=1}^{n}(2\pi\sigma_a^2 v_t)^{-1/2}\exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n}\frac{(w_t-\hat{w}_{t|t-1})^2}{v_t}\right] \tag{7.5.5}$$

这里, 量值 $\hat{w}_{t|t-1}$ 和 σ_a^2 很容易由 Kalman 滤波方法确定. 着手进行 Kalman 滤波所需初值由 r 维零值向量 $\hat{\boldsymbol{Y}}_{0|0}=0$ 和 $\boldsymbol{V}=\text{cov}[\boldsymbol{Y}_0]$ 给出. γ_k 利用第 5 章的关系式 $w_{t+j}=\hat{w}_t(j)+\sum_{k=0}^{j-1}\psi_k a_{t+j-k}$ 得出, $\boldsymbol{V}_{0|0}$ 的元素很容易确定为 w_t 的 ARMA(p,q) 过程的自协方差和权 ψ_k 的函数. 更详细的内容见 [123] . 例如, 在 w_t 的 ARMA(1,1) 模型的情形, 我们有 $\boldsymbol{Y}_t'=(w_t,\hat{w}_t(1))$, 于是

$$\boldsymbol{V}_{0|0}=\text{cov}[\boldsymbol{Y}_0]=\begin{bmatrix}\gamma_0 & \gamma_1\\ \gamma_1 & \gamma_0-\sigma_a^2\end{bmatrix}=\sigma_a^2\begin{bmatrix}\sigma_a^{-2}\gamma_0 & \sigma_a^{-2}\gamma_1\\ \sigma_a^{-1}\gamma_1 & \sigma_a^{-1}\gamma_0-1\end{bmatrix}$$

通常 ,1 期预报值 $\hat{w}_{t|t-1}$ 和相应的误差方差 $\sigma_a^2 v_t$ 相当迅速地接近于稳态 ,Kalman 滤波计算在某些阶段 (如在时间 t_0 以外) 可转化为简单形式 : $\hat{w}_{t|t-1}=\sum_{i=1}^{p}\phi_i w_{t-1}-\sum_{i=1}^{q}\theta_i a_{t-i|t-i-1}$ 及 $\sigma_a^2 v_t = \text{var}[a_{t|t-1}]=\sigma_a^2$, 对于 $t>t_0, a_{t|t-1}=w_t-\hat{w}_{t|t-1}$. 细节可见 [94] .

新息方法 用 (7.5.5) 形式所表示的似然函数通常称为新息形式, 量 $a_{t|t-1}=w_t-\hat{w}_{t|t-1}, t=1,\cdots,n$ 是 (有限样本) 新息. 以 ARMA 过程的状态空间模型表示和相关的 Kalman 滤波算法为基础, 这种新息形式似然函数的计算已由许多作者给出, 他们包括 Gardner,Harvey, 和 Phillips[94], Harvey 和 Phillips[107], 以及 Jones[123] . 似然函数的新息形式也可以不直接用状态空间模型表式得到, 而是通过使用 "新息算法 " 得到 (例如, 见 Ansley[11] 和 Brockwell 及 Davis[63]) . 这种方法主要涉及一个导出的 MA(q)

过程 $w_t' = w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$，要对该过程的 $n \times n$ 带形协方差阵做 Cholesky 分解. 更具体地，用附录 A7.3 中的记号，我们将 ARMA 模型的关系式写为 $\boldsymbol{L}_\phi \boldsymbol{w} = \boldsymbol{L}_\theta \boldsymbol{a} + \boldsymbol{F}\boldsymbol{e}_*$ 。于是，导出变量 $\boldsymbol{L}_\phi \boldsymbol{w}$ 相应向量的协方差阵是

$$\text{cov}[\boldsymbol{L}_\phi \boldsymbol{w}] = \text{cov}[\boldsymbol{L}_\theta \boldsymbol{a} + \boldsymbol{F}\boldsymbol{e}_*] = \sigma_a^2(\boldsymbol{L}_\theta \boldsymbol{L}_\theta' + \boldsymbol{F}\boldsymbol{\Omega}\boldsymbol{F}')$$

这是一个带形阵. 新息算法得到对带形阵 $\boldsymbol{L}_\theta \boldsymbol{L}\theta' + \boldsymbol{F}\boldsymbol{\Omega}\boldsymbol{F}'$ 的 (自由平方根)Cholesky 分解，记为 $\boldsymbol{G}\boldsymbol{D}\boldsymbol{G}'$, 这里， $\boldsymbol{G}$ 是对角线元为 1 的下三角带形阵， $\boldsymbol{D}$ 是具有正对角线元素 $v_t, t = 1, \cdots, n$ 的对角阵。因此，在似然函数中的二次型为

$$\begin{aligned} \boldsymbol{w}'[\text{cov}(\boldsymbol{w})]^{-1}\boldsymbol{w} &= \frac{1}{\sigma_a^2}\boldsymbol{w}'(\boldsymbol{L}_\phi^{-1}\boldsymbol{G}\boldsymbol{D}\boldsymbol{G}'\boldsymbol{L}_\phi'^{-1})^{-1}\boldsymbol{w} \\ &= \frac{1}{\sigma_a^2}\boldsymbol{e}'\boldsymbol{D}^{-1}\boldsymbol{e} = \frac{1}{\sigma_a^2}\sum_{t=1}^{n}\frac{a_{t|t-1}^2}{v_t} \end{aligned}$$

这里 $\boldsymbol{e} = \boldsymbol{G}^{-1}\boldsymbol{L}_\phi \boldsymbol{w} = (a_{1|0}, a_{2|1}, \cdots, a_{n|n-1})'$ 是新息向量，它由 $\boldsymbol{G}\boldsymbol{e} = \boldsymbol{L}_\phi \boldsymbol{w}$ 递推计算。

当序列 z_t 的某些值未被观测到, 即在 z_t 的顺序值中存在丢失值的情形, 已经证明, 有关确定似然函数的“新息”状态空间方法完全可以用于这种情况下 ARMA 模型的参数估计问题，例如可见， Jones[123],Harvey 和 Pierse[108], 以及 Wincek 和 Reinsel[205] 由递推方程 (7.5.1)–(7.5.3), 用 Kalman 滤波方法计算的精确的似然函数可通过数值最优化算法来实现极大化。这通常需要对数似然函数关于未知参数某些形式的一阶导数，利用数值导数来得到解析导数是更令人满意的. 从 (7.5.5) 中似然函数的表示形式可以看出，这涉及到对每个 $t = 1, \cdots, n$ 求一步预报 $\hat{w}_{t|t-1}$ 和误差方差 $\sigma_a^2 v_t$ 的偏导数。 Wincek 和 Reinse[205] 详细阐述了：如何通过 Kalman 滤波算法方程 (更新方程和预报方程 (7.5.1) 到 (7.5.3)) 的微分来获得 $a_{t|t-1} = w_t - \hat{w}_{t|t-1}$ 和 $\sigma_a^2 v_t = \text{var}[a_{t|t-1}]$ 关于模型参

数 $\boldsymbol{\phi},\boldsymbol{\theta}$ 和 σ_a^2 的精确导数。这样便再次导出了与似然函数 (7.5.5) 有关的极大似然估计迭代算法的显式表示，这与在 7.2 节中曾详细讨论的非线性最小二乘方法很相似。

为了说明似然函数计算的状态空间方法，我们参考在 7.1.5 节中的例子，该例中涉及来自于一个 ARMA(1,1) 模型的 $n=12$ 个数值，通过此例来说明 Kalman 滤波计算的结果，与前面一样，仍取参数值 $\phi=0.3,\theta=0.7$ 。首先，由于对于 ARMA(1,1) 模型有：$\sigma_a^{-2}\gamma_0=(1+\theta^2-2\phi\theta)/(1-\phi^2)$, $\sigma_a^{-2}\gamma_1=(1-\phi\theta)(\phi-\theta)(1-\phi^2)$, $\sigma_a^{-2}\gamma_0-1=(\phi-\theta)^2/(1-\phi^2)$, 于是我们有

$$\boldsymbol{V}_{0|0}=\sigma_a^2\begin{bmatrix}1.176 & -0.347\\ -0.347 & 0.176\end{bmatrix}$$

另外，在此例中， (7.5.1) 式中的 Kalman 增益矩阵 $\boldsymbol{K}_t$ 形式为 $\boldsymbol{K}_t=(1,k_{2t})'$, 这里 k_{2t} 等于矩阵 $\boldsymbol{V}_{t|t-1}$ 中 (2,1) 位置上的元素与 (1,1) 元素之比。通过相应于 (7.5.1) 到 (7.5.3) 的递推算式，我们求出了 $a_{t|t-1}=w_t-\hat{w}_{t|t-1}, v_t=\sigma_a^{-2}\text{var}[a_{t|t-1}]$, 和 k_{2t} 的终值，有关值列在表 7.14 中。与表 7.4 中的 $[a_t]$ 值相比较，我们特别发现， $a_{t|t-1}$ 和 v_t 基本上从 $t=7$ 之后达到了稳态形式，因为在此时刻之后 $a_{t|t-1}$ 满足 $a_{t|t-1}=w_t-0.3w_{t-1}+0.7a_{t-1|t-2}$, 并有 $v_t=1.00$, 从而 $a_{t|t-1}$ 值与表 7.4 中的 $[a_t]$ 值一致。同样地，增益 $\boldsymbol{K}_t$ 大约从 $t=7$ 之后接近于稳态值 $\boldsymbol{\Psi}=(1,\psi_1)'$, 这里 $\psi_1=\phi-\theta=-0.4$, 这意味着 (7.5.1) 中的方程退化为稳态形式 $\hat{w}_{t|t-1}=0.3w_{t-1}-0.7a_{t-1|t-2}$ 。另外，我们求出用 Kalman 滤波算法得到的精确平方和值为 $S(0.3,0.7)=\sum_{t=1}^{12}a_{t|t-1}^2/v_t=89.158$, 这与在 7.1.5 节中用不同方法求得的结果一样， 同时得到行列式值 $\prod_{t=1}^{12}v_t=|\boldsymbol{M}_n^{(1,1)}|^{-1}=1.345$, 这与由下面关系式

$$|\boldsymbol{M}_n^{(1,1)}|^{-1}=1+\frac{1-\theta^{2n}}{1-\theta^2}\frac{(\phi-\theta)^2}{1-\phi^2}$$

所求得的值一样。该式在附录 A7.3 中用另一种方法导出。

表 7.14　对于由模型 $(1-0.3B)w_t=(1-0.7B)a_t$ 产生的 12 个值用 Kalman 滤波的递推计算

t	$a_{t\|t-1}$	v_t	k_{2t}	t	$a_{t\|t-1}$	v_t	k_{2t}
1	1.10	1.18	−0.30	7	2.65	1.00	−0.40
2	4.62	1.07	−0.35	8	−0.14	1.00	−0.40
3	4.73	1.03	−0.38	9	0.17	1.00	−0.40
4	1.60	1.02	−0.39	10	−0.09	1.00	−0.40
5	2.91	1.01	−0.39	11	0.83	1.00	−0.40
6	4.74	1.00	−0.40	12	2.34	1.00	−0.40

A 7.1　正态分布理论的回顾

A7.1.1　正定二次型的分块

考虑正定二次型 $Q_p=\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}$ 。假设 $p\times1$ 向量 $\boldsymbol{x}$ 在第 p_1 个元素之后分块，即 $\boldsymbol{x}'=(\boldsymbol{x}_1'\vdots\boldsymbol{x}_2')=(x_1,x_2,\cdots,x_{p_1}\vdots x_{p_1+1},\cdots,x_p)$，并假设 $p\times p$ 矩阵 $\boldsymbol{\Sigma}$ 也是在第 p_1 行和第 p_2 列后分块，于是

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \vdots & \boldsymbol{\Sigma}_{12}\\ \cdots & \cdots & \cdots\\ \boldsymbol{\Sigma}_{12'} & \vdots & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

那么，由于

$$\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}=(\boldsymbol{x}_1'\vdots\boldsymbol{x}_2')$$

$$\times\begin{bmatrix}\boldsymbol{I} & \vdots & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\\ \cdots & \cdots & \cdots\\ \boldsymbol{0} & \vdots & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{11}^{-1} & \vdots & \boldsymbol{0}\\ \cdots & \cdots & \cdots\\ \boldsymbol{0} & \vdots & (\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \vdots & \boldsymbol{0}\\ \cdots & \cdots & \cdots\\ -\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1} & \vdots & \boldsymbol{I}\end{bmatrix}\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}$$

$Q_p=\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}$ 总可以写成两个二次型 Q_{p_1} 和 Q_{p_2} 之和，它们分别包含 p_1 和 p_2 个元素，并有

$$\begin{aligned}Q_p&=Q_{p_1}+Q_{p_2}\\ Q_{p_1}&=\boldsymbol{x}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{x}_1\\ Q_{p_2}&=(\boldsymbol{x}_2-\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{x}_1)'(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}(\boldsymbol{x}_2-\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{x}_1)\end{aligned}\tag{A7.1.1}$$

我们也可以将 $\boldsymbol{\Sigma}$ 的行列式写为

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}| \tag{A7.1.2}$$

A7.1.2 两个有用的积分

设 $\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}$ 是 $\boldsymbol{z}$ 的一个正定二次型， $\boldsymbol{z}$ 有 q 个元素，即 $\boldsymbol{z}' = (z_1, z_2, \cdots, z_q)$，其中 $-\infty < z_i < +\infty, i = 1, 2, \cdots, q$，设 a, b 和 m 是正实数。则可证明有

$$\int_R \left(a + \frac{\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}}{b}\right)^{-(m+q)/2} d\boldsymbol{z} = \frac{(b\pi)^{q/2}\Gamma(m/2)}{a^{m/2}|\boldsymbol{C}|^{1/2}\Gamma[(m+q)/2]} \tag{A7.1.3}$$

这里 q 重积分扩展到 $\boldsymbol{z}$ 的全空间 R 上，且有

$$\frac{\int_{\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}>qF_0}(1 + \boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}/m)^{-(m+q)/2}d\boldsymbol{z}}{\int_{\mathrm{R}}(1 + \boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}/m)^{-(m+q)/2}d\boldsymbol{z}} = \int_{F_0}^{\infty} p(F|q, m)dF \tag{A7.1.4}$$

其中函数 $p(F|q, m)$ 称为具有自由度 q 和 m 的 F 分布，且定义为

$$p(F|q, m) = \frac{(q/m)^{q/2}\Gamma[(m+q)/2]}{\Gamma(q/2)\Gamma(m/2)} F^{(q-2)/2}\left(1 + \frac{q}{m}F\right)^{-(m+q)/2} \quad (F > 0) \tag{A7.1.5}$$

若 m 趋于无穷，则

$$\left(1 + \frac{\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}}{m}\right)^{-(m+q)/2} \quad \text{趋于 } e^{-(\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z})/2}$$

记 $qF = \chi^2$，我们由 (A7.1.4) 得到

$$\frac{\int_{\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}>\chi_0^2} e^{-(\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z})/2}d\boldsymbol{z}}{\int_R e^{-(\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z})/2}d\boldsymbol{z}} = \int_{\chi_0^2}^{\infty} p(\chi^2|q)d\chi^2 \tag{A7.1.6}$$

这里，函数 $p(\chi^2|q)$ 称为自由度为 q 的 χ^2 分布，它定义为

$$p(\chi^2|q) = \frac{1}{2^{q/2}\Gamma(q/2)}(\chi^2)^{(q-2)/2}e^{-\chi^2/2} \quad (\chi^2 > 0) \qquad \text{(A7.1.7)}$$

这里及别处都用 $p(x)$ 作为一般记号来表示一个随机变量 x 的概率密度函数.

A7.1.3 正态分布

若随机变量 x 的概率密度为

$$p(x) = (2\pi)^{1/2}(\sigma^2)^{-1/2}e^{-(x-\mu)^2/2\sigma^2} \qquad \text{(A7.1.8)}$$

则称该随机变量为具有均值 μ 、方差 σ 的正态分布，或称其分布为 $N(\mu,\sigma^2)$. 因此，单位正态偏差 $u = (x-\mu)/\sigma$ 具有分布 $N(0,1)$. 在第五部分的表 E 给出了纵坐标 $p(u = u_\epsilon)$ 和值 u_ϵ, 对于给定 ϵ , 它们满足 $\Pr\{u > u_\epsilon\} = \epsilon$.

多维正态分布　向量随机变量 $\boldsymbol{x}' = (x_1, x_2, \cdots, x_2)$ 称为具有 p- 变量的联合正态分布 $N\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, 若它的概率密度函数为

$$p(\boldsymbol{x}) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2} \qquad \text{(A7.1.9)}$$

概率密度等值线是由 $(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ 等于常数所定义的椭圆. 作为例子，图 A7.1 中给出了两变量正态分布的椭圆等值线. 在点 $\boldsymbol{x} = \boldsymbol{\mu}$ 处，多维分布具有最大的密度值

$$\max p(\boldsymbol{x}) = p(\boldsymbol{\mu}) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}$$

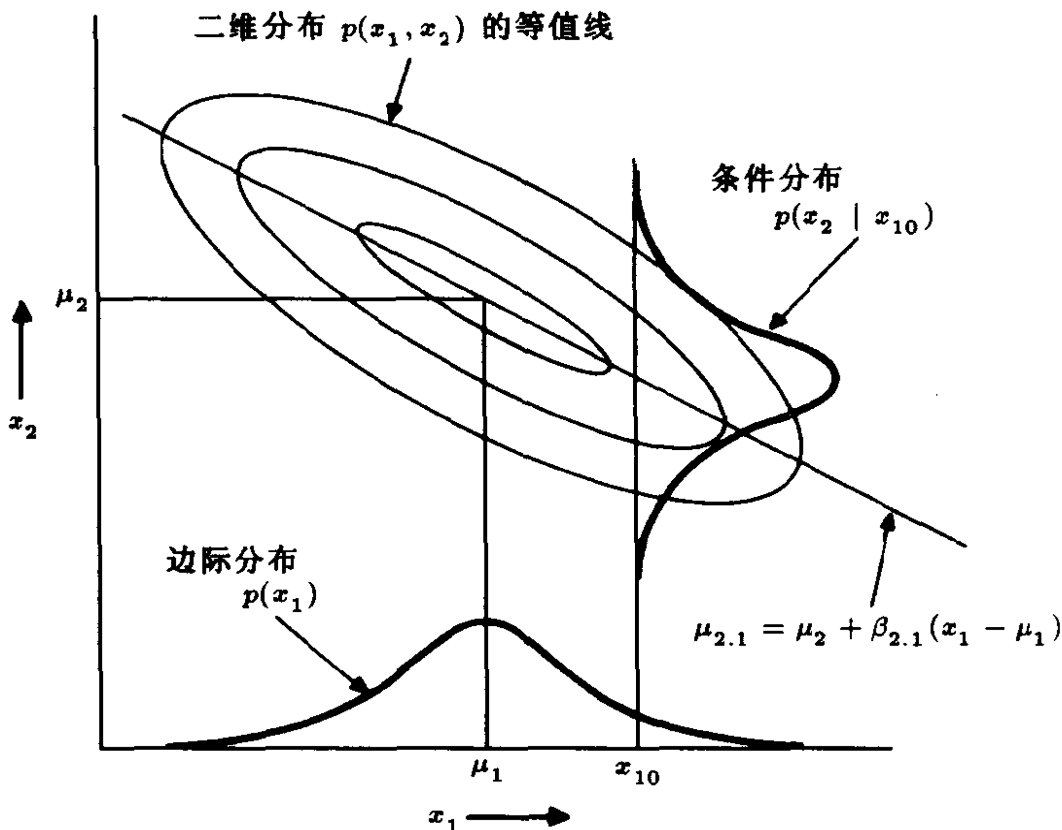


图 A7.1 二维分布的等值线，同时给出边际分布 $p(x_t)$ 和在 $x_1 = x_{10}$ 处的条件分布 $p(x_2|x_{10})$

χ^2 分布是作为多维正态密度等值线以外的概率总和　对于 p- 变量正态分布 (A7.1.9), 其密度等值线定义为

$$(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \chi_0^2$$

在该等值线以外的概率总和定义为下面的 p 自由度的 χ^2 积分:

$$\int_{\chi_0^2}^{\infty} p(\chi^2|p)d\chi^2$$

其中 χ^2 的密度函数定义如 (A7.1.7)．在第五部分的表 F 中 $\chi_\epsilon^2(p)$ 的值，对于任意给定的 ϵ 有 $\Pr\{\chi^2 > \chi_\epsilon^2(p)\} = \epsilon$ ．

多维正态的边际分布和条件分布　设 $p = p_1 + p_2$ 变量的随机向量在前 p_1 个元素之后被分块，即

$$\boldsymbol{x}' = (\boldsymbol{x}_1' : \boldsymbol{x}_2') = (x_1, x_2, \cdots, x_{p_1} : x_{p_1+1}, \cdots, x_{p_1+p_2})$$

其方差 - 协方差阵为

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \vdots & \boldsymbol{\Sigma}_{12}\\ \cdots & \cdots & \cdots\\ \boldsymbol{\Sigma}_{12}' & \vdots & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

于是由 (A7.1.2) 和 (A7.1.2) 我们可以把 $p=p_1+p_2$ 变量的多维正态分布写为 $\boldsymbol{x}_1$ 的边际分布与给定 $\boldsymbol{x}_1$ 时 $\boldsymbol{x}_2$ 的条件分布的乘积，即

$$\begin{aligned}p(\boldsymbol{x})&=p(\boldsymbol{x}_1,\boldsymbol{x}_2)=p(\boldsymbol{x}_1)p(\boldsymbol{x}_2|\boldsymbol{x}_1)\\&=(2\pi)^{-p_1/2}|\boldsymbol{\Sigma}_{11}|^{-1/2}\exp\left[-\frac{(\boldsymbol{x}_1-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)}{2}\right]\\&\quad\times(2\pi)^{-p_2/2}|\boldsymbol{\Sigma}_{22.11}|^{-1/2}\exp\left[-\frac{(\boldsymbol{x}_2-\boldsymbol{\mu}_{2.1})'\boldsymbol{\Sigma}_{22.11}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_{2.1})}{2}\right]\end{aligned} \tag{A7.1.10}$$

其中

$$\boldsymbol{\Sigma}_{22.11}=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \tag{A7.1.11}$$

且 $\boldsymbol{\mu}_{2.1}=\boldsymbol{\mu}_2+\boldsymbol{\beta}_{2.1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)=E[\boldsymbol{x}_2|\boldsymbol{x}_1]$ 定义了在 (p_1+p_2) 空间中的回归超平面，它刻画了当 $\boldsymbol{x}_1$ 的 p_1 个元素变化时，$\boldsymbol{x}_2$ 的 p_2 个元素的平均轨迹. 回归系数的 $p_1\times p_2$ 阵由 $\boldsymbol{\beta}_{2.1}=\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{11}^{-1}$ 给出。

由此可知，多维正态的边际分布和条件分布都是多维正态分布。可以看出，对于多维正态来说，除了其位置 (即均值) 以外，不管 $\boldsymbol{x}_1$ 是什么值，条件分布 $p(\boldsymbol{x}_2|\boldsymbol{x}_1)$ 都是相同的。

单变量边际分布　特别地，对于单个元素 $x_i(i=1,2,\cdots,p)$ 的边际密度就是 $N(\mu_i,\sigma_i^2)$, 这是一个单变量正态分布，其均值等于 $\boldsymbol{\mu}$ 的第 i 元素，方差是 $\boldsymbol{\Sigma}$ 的第 i 个对角线元素.

二维正态　为了说明起见，图 A7.1 给出了一个二维正态的边际分布和条件分布. 在此例中，x_1 的边际分布是 $N(\mu_1,\sigma_1^2)$, 而给定 x_1 时 x_2 的条件分布为

$$N\left\{\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1),\sigma_2^2(1-\rho^2)\right\}$$

其中 $\rho = (\sigma_1/\sigma_2)\beta_{2.1}$ 是 x_1 和 x_2 之间的相关系数.

A7.1.4 学生 t– 分布

若随机变量 x 的分布密度为

$$p(x) = (2\pi)^{-1/2}(s^2)^{-1/2}\left(\frac{\nu}{2}\right)^{-1/2}\Gamma\left(\frac{\nu+1}{2}\right) \times \Gamma^{-1}\left(\frac{\nu}{2}\right)\left[1+\frac{(x-\mu)^2}{\nu s^2}\right]^{-(\nu+1)/2} \tag{A7.1.12}$$

则称其为具有标度的 t 分布 $t(\mu, s^2, \nu)$, 其中 μ 为均值, s 是标度参数, ν 是自由度.

因此, 将 t 标准化的偏离 $t = (x-\mu)/s$ 具有分布 $t(0,1,\nu)$. 在第五部分的表 G 中给出了 t_ϵ 的值, 对于选定值 ϵ 有 $\Pr\{t > t_\epsilon\} = \epsilon$.

趋近于正态　对于很大的 ν, 下面乘积

$$\left(\frac{\nu}{2}\right)^{-1/2}\Gamma\left(\frac{\nu+1}{2}\right)\Gamma^{-1}\left(\frac{\nu}{2}\right)$$

趋于 1, 而 (A7.1.12) 右边的方括号项趋于 $e^{(-1/2s^2)(x-\mu)^2}$. 于是, 对于大的 ν 值, 我们可记 $s^2 = \sigma^2$, 因此, t 分布趋于正态分布 (A7.1.8).

多维 t 分布　设 $\boldsymbol{\mu}'$ 为一 $p \times 1$ 向量 $(\mu_1, \mu_2, \cdots, \mu_p)$, $\boldsymbol{S}$ 是 $p \times p$ 正定阵. 若向量随机变量 $\boldsymbol{x}$ 的分布密度为

$$p(\boldsymbol{x}) = (2\pi)^{-p/2}|\boldsymbol{S}|^{-1/2}\left(\frac{\nu}{2}\right)^{-p/2}\Gamma\left(\frac{\nu+p}{2}\right)\Gamma^{-1}\left(\frac{\nu}{2}\right) \times \left[1+\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right]^{-(\nu+p)/2} \tag{A7.1.13}$$

则称 $\boldsymbol{x}$ 为具有标量 t 分布 $t(\boldsymbol{\mu}, \boldsymbol{S}, \nu)$(如 [73],[86] 所述), 其中, $\boldsymbol{\mu}$ 为均值, $\boldsymbol{S}$ 是标度矩阵, ν 为自由度. 多维 t 分布的等值线是 $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=$ 常数所定义的椭圆.

趋近于多维正态形式　对于很大的 ν, 下面乘积

$$\left(\frac{\nu}{2}\right)^{-p/2}\Gamma\left(\frac{\nu+1}{2}\right)\Gamma^{-1}\left(\frac{\nu}{2}\right)$$

趋于 1, 而 (A7.1.13) 右边的方括号项趋于 $e^{-(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2}$. 于是，对于大的 ν 值，我们可记 $\boldsymbol{S}=\boldsymbol{\Sigma}$, 因此，多维 t 分布趋于多维正态分布 (A7.1.9) .

F 分布作为多维 t 分布密度等值线以外的概率总和　由 (A7.1.4), p 维 t 分布 $t(\boldsymbol{\mu},\boldsymbol{S},\nu)$ 的密度等值线定义为

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=pF_0$$

在该等值线以外的概率总和由下面具有自由度 p 和 ν 的 F 积分给出 $\int_{F_0}^{\infty}p(F|p,\nu)dF$ 这里，关于 F 的密度函数由 (A7.1.5) 定义。对于大的 ν,$\Pr\{F>F_\epsilon(p,\nu)\}=\Pr\{\chi^2>\chi^2_\epsilon(p)\}$, 其中 $pF=\chi^2$. 因此，可以料想，当 ν 很大时，多维 t 分布密度等值线以外的总和等于多维 t 分布所逼近的多维正态分布相应密度等值线以外的总和。

边际 t 分布　若 $p=p_1+p_2$ 维向量 $\boldsymbol{x}$, 其分布如方程 (A7.1.13) 给出，将 $\boldsymbol{x}$ 从第 p_1 元素后分块，即 $\boldsymbol{x}'=(\boldsymbol{x}_1' : \boldsymbol{x}_2')$, 那么，对 $\boldsymbol{S}$ 作类似分块，我们得到

$$\boldsymbol{S}=\begin{bmatrix}\boldsymbol{S}_{11} & \vdots & \boldsymbol{S}_{12}\\ \cdots & \cdots & \cdots\\ \boldsymbol{S}_{12}' & \vdots & \boldsymbol{S}_{22}\end{bmatrix}$$

记

$$\boldsymbol{S}_{22.11}=\boldsymbol{S}_{22}-\boldsymbol{S}_{12}'\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}$$
$$\boldsymbol{\mu}_{2.1}=\boldsymbol{\mu}_2+\boldsymbol{\beta}_{2.1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)\qquad \boldsymbol{\beta}_{2.1}=\boldsymbol{S}_{12}'\boldsymbol{S}_{11}^{-1}$$

那么有

$$p(\boldsymbol{x}_1,\boldsymbol{x}_2)=(2\pi)^{-(p_1+p_2)/2}|\boldsymbol{S}_{11}|^{-1/2}|\boldsymbol{S}_{22.11}|^{-1/2}\left(\frac{\nu}{2}\right)^{-(p_1+p_2)/2}$$
$$\times\Gamma\left(\frac{\nu+p_1+p_2}{2}\right)\Gamma^{-1}\left(\frac{\nu}{2}\right)$$
$$\times\left[1+\frac{(\boldsymbol{x}_1-\boldsymbol{\mu}_1)'\boldsymbol{S}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)}{\nu}+\frac{(\boldsymbol{x}_2-\boldsymbol{\mu}_{2.1})'\boldsymbol{S}_{22.11}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_{2.1})}{\nu}\right]^{-(\nu+p_1+p_2)/2}$$

现在利用先前的结果 (A7.1.3), 有

$$a=1+\frac{(\boldsymbol{x}_1-\boldsymbol{\mu}_1)'\boldsymbol{S}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)}{\nu}$$

$b=\nu, m=\nu+p_1, q=p_2, \boldsymbol{C}=\boldsymbol{S}_{22.11}^{-1}$, 我们得到

$$p(\boldsymbol{x}_1)=(2\pi)^{-p_1/2}|\boldsymbol{S}_{11}|^{-1/2}\left(\frac{\nu}{2}\right)^{-p1/2}\Gamma\left(\frac{\nu+p_1}{2}\right)\Gamma^{-1}\left(\frac{\nu}{2}\right)$$
$$\times\left[1+\frac{(\boldsymbol{x}_1-\boldsymbol{\mu}_1)'\boldsymbol{S}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{\mu}_1)}{\nu}\right]^{-(\nu+p_1)/2} \tag{A7.1.14}$$

因此, 若一个 p 维向量 $\boldsymbol{x}$ 具有等式 (A7.1.13) 的多维 t 分布, 则任意 p_1 个变量 $\boldsymbol{x}_1$ 的分布是 p_1- 变量 t 分布 $t(\boldsymbol{\mu}_1,\boldsymbol{S}_{11},\nu)$ 。

单变量 t 分布　特别地, 仅一个元素 $x_i(i=1,2,\cdots,p)$ 的边际分布为 $t(\mu_i,s_{ii},\nu)$, 这是一个一元 t 分布, 均值为 $\boldsymbol{\mu}$ 的第 i 个元素, 标度因子等于 $\boldsymbol{S}$ 第 i 个对角线元素的正平方根, 自由度为 ν 。

条件 t 分布　条件分布 $p(\boldsymbol{x}_2|\boldsymbol{x}_1)$ 可由联合分布 $p(\boldsymbol{x}_1,\boldsymbol{x}_2)$ 与边际分布 $p(\boldsymbol{x}_1)$ 之比得到, 于是

$$p(\boldsymbol{x}_2|\boldsymbol{x}_{10})=\frac{p(\boldsymbol{x}_{10},\boldsymbol{x}_2)}{p(\boldsymbol{x}_{10})}$$

经分解后, 我们有

$$p(\boldsymbol{x}_2|\boldsymbol{x}_{10})$$
$$=\text{常数}\left\{1+\frac{(\boldsymbol{x}_2-\boldsymbol{\mu}_{2.1})'[c(\boldsymbol{x}_{10})\boldsymbol{S}_{22.11}^{-1}](\boldsymbol{x}_2-\boldsymbol{\mu}_{2.1})}{\nu+p_1}\right\}^{-(\nu+p_1+p_2)/2} \tag{A7.1.15}$$

其中 $c(\boldsymbol{x}_{10})$(对于给定 $\boldsymbol{x}_{10}$) 由下式给出

$$c(\boldsymbol{x}_{10}) = \frac{\nu + p_1}{\nu + (\boldsymbol{x}_{10} - \boldsymbol{\mu}_1)'\boldsymbol{S}_{11}^{-1}(\boldsymbol{x}_{10} - \boldsymbol{\mu}_1)} \tag{A7.1.16}$$

因此，对于给定的 $\boldsymbol{x}_1 = \boldsymbol{x}_{10}, \boldsymbol{x}_2$ 的分布是多维 t 分布

$$t\{\boldsymbol{\mu}_2 + \boldsymbol{\beta}_{2.1}(\boldsymbol{x}_{10} - \mu_1), c^{-1}(\boldsymbol{x}_{10})\boldsymbol{S}_{22.11}, \nu + p_1\}$$

这里 $\boldsymbol{\beta}_{22.11} = \boldsymbol{S}_{12}'\boldsymbol{S}_{11}^{-1}$

二维 t 分布　如前所述，通过研究二维分布 $t(\boldsymbol{\mu}, \boldsymbol{S}, \nu)$ 可以得出对于一般多维情形的一些了解。其图形显示类似于图 A7.1, 看上去和二维正态很相象。然而相应边际分布是具有自由度 ν 的单变量 t 分布，而条件分布却是自由度为 $\nu + 1$ 的 t 分布。而且，条件分布比如 $p(x_2|x_{10})$ 的标度因子将依赖于 x_{10}, 这与正态情形是不同的，正态的条件分布具有与 x_{10} 无关的相同的方差。

A 7.2　线性最小二乘原理的回顾

A7.2.1　正规方程

假设模型为

$$w_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \tag{A7.2.1}$$

这里 $w_i (i = 1, 2, \cdots, n)$ 是从某个实验得到的观测值，其中独立变量 $x_{i1}, x_{i2}, \cdots, x_{ik}$ 取确定的已知量，β_i 是有待于从数据估计的未知参数，e_i 是不相关误差，其均值为 0, 且具有相同方差 σ^2 。

关系式 (A7.21) 可以改写成矩阵表示

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

或

$$\boldsymbol{w} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{A7.2.2}$$

这里 $\boldsymbol{X}$ 被假定为是 k 满秩. 在 [23] 中叙述的 Gauss 最小二乘原理可表述如下：参数 $\boldsymbol{\beta}$ 的估计值 $\hat{\boldsymbol{\beta}}' = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k)$ 对于观测值是线性的且关于 $\boldsymbol{\beta}$ 是无偏的，它在所有这样的函数中使得参数的线性函数 $\lambda_1\beta_1 + \lambda_2\beta_2 + \cdots + \lambda_k\beta_k$ 的均方误差最小，它是通过极小化如下平方和求得：

$$S(\boldsymbol{\beta}) = \boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{w} - \boldsymbol{X\beta})'(\boldsymbol{w} - \boldsymbol{X\beta}) \tag{A7.2.3}$$

为了得到 $S(\boldsymbol{\beta})$ 的最小值，我们注意到向量 $\boldsymbol{w} - \boldsymbol{X\beta}$ 可以分解为两个向量 $\boldsymbol{w} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 和 $\boldsymbol{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$, 即

$$\boldsymbol{w} - \boldsymbol{X\beta} = \boldsymbol{w} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \tag{A7.2.4}$$

因此得出，若我们选定 $\hat{\boldsymbol{\beta}}$ 使得

$$(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{w} \tag{A7.2.5}$$

则有

$$S(\boldsymbol{\beta}) = S(\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \tag{A7.2.6}$$

这里向量 $\boldsymbol{w} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ 和 $\boldsymbol{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$ 是正交的。由于 (A7.2.6) 中的第二项是一个正定二次型，所以当 $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ 时可求得最小值，这里 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{w}$ 由正规方程 (A7.2.5) 给出。

A7.2.2 残差方差的估计

利用 (A7.2.3) 和 (A7.2.5), 在最小值处的平方和为

$$S(\hat{\boldsymbol{\beta}}) = \boldsymbol{w}'\boldsymbol{w} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{A7.2.7}$$

从而，若我们定义

$$s^2 = \frac{S(\hat{\boldsymbol{\beta}})}{n-k} \tag{A7.2.8}$$

则可以证明 [158]: $E[s^2]=\sigma^2$, 因此, s^2 给出了 σ^2 的一个无偏估计。

A7.2.3 估计值的协方差阵

由于 $\mathrm{cov}[\boldsymbol{w},\boldsymbol{w}']=\boldsymbol{I}\sigma^2$, 估计值的协方差阵可定义为

$$\begin{aligned}V(\hat{\boldsymbol{\beta}})&=\mathrm{cov}[\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\beta}}]\\&=\mathrm{cov}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{w},\boldsymbol{w}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]\\&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2\end{aligned}\tag{A7.2.9}$$

A7.2.4 置信域

在正态假设下 [158],(A7.2.6) 中的二次型 $S(\hat{\boldsymbol{\beta}})$ 和 $(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})$ 是独立的, 具有自由度分别为 $n-k$ 和 k 的 σ^2 倍 χ^2 随机变量的分布。因此,

$$\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{S(\hat{\boldsymbol{\beta}})}\,\frac{n-k}{k}$$

分布为 $F(k,n-k)$ 。利用 (A7.2.8) 可得到

$$(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})<ks^2F_\epsilon(k,n-k)\tag{A7.2.10}$$

定义了 $\boldsymbol{\beta}$ 的 $1-\epsilon$ 置信域。

A7.2.5 相关误差

假设 (A7.2.2) 中的误差 $\boldsymbol{e}$ 有已知的方差阵 $\boldsymbol{V}$, 这里 $\boldsymbol{V}^{-1}=\boldsymbol{P}\boldsymbol{P}'/\sigma^2$ 。则 (7.2.2) 可以写为

$$\boldsymbol{P}'\boldsymbol{w}=\boldsymbol{P}'\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{P}'\boldsymbol{e}$$

或

$$\boldsymbol{w}^*=\boldsymbol{X}^*\boldsymbol{\beta}+\boldsymbol{e}^*\tag{A7.2.11}$$

$\boldsymbol{e}^* = \boldsymbol{P}'\boldsymbol{e}$ 的协方差阵是

$$\text{cov}[\boldsymbol{P}'\boldsymbol{e}, \boldsymbol{e}'\boldsymbol{P}] = \boldsymbol{P}'\boldsymbol{V}\boldsymbol{P} = \boldsymbol{I}\sigma^2$$

因此我们可以对变换后的模型 (A7.2.11) 令 $\boldsymbol{V} = \boldsymbol{I}\sigma^2$ 应用一般最小二乘原理。在这里， $\boldsymbol{w}$ 由 $\boldsymbol{w}^* = \boldsymbol{P}'\boldsymbol{w}$ 代替， $\boldsymbol{X}$ 由 $\boldsymbol{X}^* = \boldsymbol{P}'\boldsymbol{X}$ 代替。

A 7.3 滑动平均和混合过程的精确似然函数

为了对 MA(q) 模型求得所需要的似然函数，我们必须对 $\boldsymbol{w}' = (w_1, w_2, \cdots, w_n)$ 推导概率密度函数，假设该序列是由下面的 q 阶滑动平均模型产生：

$$\widetilde{w}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \qquad \text{(A7.3.1)}$$

其中 $\hat{w}_t = w_t - \mu, \mu = E[w_t]$ 。在 a 为正态，从而 $\widetilde{w}$ 也为正态分布的假设下，联合密度可以写为

$$p(\boldsymbol{w}|\boldsymbol{\theta}, \sigma_a, \mu) = (2\pi\sigma_a^2)|\boldsymbol{M}_n^{(0,q)}|^{1/2} \exp\left[\frac{-\widetilde{\boldsymbol{w}}'\boldsymbol{M}_n^{(0,q)}\widetilde{\boldsymbol{w}}}{2\sigma_a^2}\right] \qquad \text{(A7.3.2)}$$

这里，$(\boldsymbol{M}_n^{(p,q)})^{-1}\sigma_a^2$ 记 ARMA(p,q) 过程中 w 的 $n \times n$ 协方差阵。我们现在考虑一种计算 $\widetilde{\boldsymbol{w}}'\boldsymbol{M}_n^{(0,q)}\widetilde{\boldsymbol{w}}$ 的方便方法，为简便起见，我们假设 $\mu = 0$, 于是 $w_t = \widetilde{w}_t$ 。

由模型 (A7.3.1) 我们可以写出 n 个方程

$$w_t = q_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (t = 1, 2, \cdots, n)$$

这 n 个方程可以用 n 维向量 $\boldsymbol{w}' = (w_1, w_2, \cdots, w_n)$,$\boldsymbol{a}' = (a_1, a_2, \cdots, a_n)$, 及 q 维初值向量 $\boldsymbol{a}'_* = (a_{1-q}, a_{2-q}, \cdots, a_0)$ 用矩阵形式表示为

$$\boldsymbol{w} = \boldsymbol{L}_\theta \boldsymbol{a} + \boldsymbol{F}\boldsymbol{a}_*$$

这里 $\boldsymbol{L}_\theta$ 是主对角线元素为 1 的 $n\times n$ 下三角阵，第一次角线为 $-\theta_1$, 第二次角线为 $-\theta_2$, 如此等等，当 $i>q$ 时有 $\theta_i=0$ 。进而，$\boldsymbol{F}$ 是 $n\times q$ 阵，有 $\boldsymbol{F}=(\boldsymbol{B}'_q,\boldsymbol{0}')'$, 这里 $\boldsymbol{B}_q$ 是 $q\times q$ 阵且等于

$$\boldsymbol{B}_q=-\begin{bmatrix}\theta_q & \theta_{q-1} & \cdots & \theta_1\\ 0 & \theta_q & \cdots & \theta_2\\ \cdots & \cdots & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & \theta_q\end{bmatrix}$$

现在，$(\boldsymbol{a}',\boldsymbol{a}'_*)$ 中 $n+q$ 个元素的联合分布为

$$p(\boldsymbol{a},\boldsymbol{a}_*|\sigma_a)=(2\pi\sigma_a^2)^{-(n+q)/2}\exp\left[-\frac{1}{2\sigma_a^2}(\boldsymbol{a}'\boldsymbol{a}+\boldsymbol{a}'_*\boldsymbol{a}_*)\right]$$

我们注意到，从 $(\boldsymbol{a},\boldsymbol{a}_*)$ 到 $(\boldsymbol{w},\boldsymbol{a}_*)$ 的变换是单位 Jacobi 型且 $\boldsymbol{a}=\boldsymbol{L}_\theta^{-1}(\boldsymbol{w}-\boldsymbol{F}\boldsymbol{a}_*)$, $\boldsymbol{w}=\boldsymbol{L}_\theta\boldsymbol{a}+\boldsymbol{F}\boldsymbol{a}_*$ 和 $\boldsymbol{a}_*$ 的联合分布是

$$p(\boldsymbol{w},\boldsymbol{a}_*|\boldsymbol{\theta},\sigma_a)=(2\pi\sigma_a^2)^{-(n+q)/2}\exp\left[-\frac{1}{2\sigma_a^2}S(\boldsymbol{\theta},\boldsymbol{a}_*)\right]$$

这里

$$S(\boldsymbol{\theta},\boldsymbol{a}_*)=(\boldsymbol{w}-\boldsymbol{F}\boldsymbol{a}_*)'\boldsymbol{L}_\theta^{'-1}\boldsymbol{L}_\theta^{-1}(\boldsymbol{w}-\boldsymbol{F}\boldsymbol{a}_*)+\boldsymbol{a}'_*\boldsymbol{a}_* \qquad \text{(A7.3.3)}$$

现在设 $\hat{\boldsymbol{a}}_*$ 是使达 $S(\boldsymbol{\theta},\boldsymbol{a}_*)$ 极小值的向量，从一般的最小二乘原理可以证明有 $\hat{\boldsymbol{a}}_*=\boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{L}_\theta^{'-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{w}$, 其中 $\boldsymbol{D}=\boldsymbol{I}_q+\boldsymbol{F}'\boldsymbol{L}_\theta^{'-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{F}$ 。因此利用 (A7.2.6) 的结果，我们有

$$S(\boldsymbol{\theta},\boldsymbol{a}_*)=S(\boldsymbol{\theta})+(\boldsymbol{a}_*-\hat{\boldsymbol{a}}_*)'\boldsymbol{D}(\boldsymbol{a}_*-\hat{\boldsymbol{a}}_*)$$

其中

$$S(\boldsymbol{\theta})=S(\boldsymbol{\theta},\hat{\boldsymbol{a}}_*)=(\boldsymbol{w}-\boldsymbol{F}\hat{\boldsymbol{a}}_*)'\boldsymbol{L}_\theta^{'-1}\boldsymbol{L}_\theta^{-1}(\boldsymbol{w}-\boldsymbol{F}\hat{\boldsymbol{a}}_*)+\hat{\boldsymbol{a}}'_*\hat{\boldsymbol{a}}_* \qquad \text{(A7.3.4)}$$

是观测值而不是初始值 $\boldsymbol{a}_*$ 的函数，于是

$$\begin{aligned}&p(\boldsymbol{w},\boldsymbol{a}_*|\boldsymbol{\theta},\sigma_a)\\&=(2\pi\sigma_a^2)^{-(n+q)/2}\exp\left\{-\frac{1}{2\sigma_a^2}\left[S(\boldsymbol{\theta})+(\boldsymbol{a}_*-\hat{\boldsymbol{a}}_*)'\boldsymbol{D}(\boldsymbol{a}_*-\hat{\boldsymbol{a}}_*)\right]\right\}\end{aligned}$$

然而，由于 $\boldsymbol{w}$ 和 $\boldsymbol{a}_*$ 的联合分布可因子分解为

$$p(\boldsymbol{w},\boldsymbol{a}_*|\boldsymbol{\theta},\sigma_a)=p(\boldsymbol{w}|\boldsymbol{\theta},\sigma_a)p(\boldsymbol{a}_*|\boldsymbol{w},\boldsymbol{\theta},\sigma_a)$$

因此，与 (A7.1.10) 类似，有

$$\begin{aligned}p(\boldsymbol{a}_*|\boldsymbol{w},\boldsymbol{\theta},\sigma_a)&=(2\pi\sigma_a^2)^{-q/2}|\boldsymbol{D}|^{1/2}\\&\quad\times\exp\left[-\frac{1}{2\sigma_a^2}(\boldsymbol{a}_*-\hat{\boldsymbol{a}}_*)'\boldsymbol{D}(\boldsymbol{a}_*-\hat{\boldsymbol{a}}_*)\right]\end{aligned}\tag{A7.3.5}$$

$$p(\boldsymbol{w}|\boldsymbol{\theta},\sigma_a)=(2\pi\sigma_a^2)^{-n/2}|\boldsymbol{D}|^{-1/2}\exp\left[-\frac{1}{2\sigma_a^2}S(\boldsymbol{\theta})\right]\tag{A7.3.6}$$

现在我们可以推导出以下结果：

1. 由 (A7.3.5) 我们看到，$\hat{\boldsymbol{a}}_*$ 是在给定 $\boldsymbol{w}$ 和 $\boldsymbol{\theta}$ 时 $\boldsymbol{a}_*$ 的条件期望。因此，利用在 7.1.4 节中引入的记号，我们得到

$$\hat{\boldsymbol{a}}_*=[\boldsymbol{a}_*|\boldsymbol{w},\boldsymbol{\theta}]=[\boldsymbol{a}_*]$$

这是 $[\boldsymbol{a}]=\boldsymbol{L}_\theta^{-1}(\boldsymbol{w}-\boldsymbol{F}[\boldsymbol{a}_*])$ 是给定 $\boldsymbol{z}$ 和 $\boldsymbol{w}$ 时 $\boldsymbol{a}$ 的条件期望，利用 (A7.3.4) 有

$$S(\boldsymbol{\theta})=[\boldsymbol{a}'][\boldsymbol{a}]+[\boldsymbol{a}_*]'[\boldsymbol{a}_*]=\sum_{t=1-q}^{n}[a_t]^2\tag{A7.3.7}$$

为了计算 $S(\boldsymbol{\theta})$，量值 $[a_t]=[a_t|\boldsymbol{w},\boldsymbol{\theta}]$ 可以利用估计值 $[\boldsymbol{a}_*]'=([a_{1-q}],[a_{2-q}],\cdots,[a_0])$，使用和前面一样的对初值反向预报的方法得出，并且从关系式 $\boldsymbol{L}_\theta\boldsymbol{a}=\boldsymbol{w}-\boldsymbol{F}[\boldsymbol{a}_*]$，可以由递推计算而获得 $[\boldsymbol{a}]$ 的元素 $[a_1],[a_2],\cdots,[a_n]$ 为

$$[a_t]=w_t+\theta_1[a_{t-1}]+\cdots+\theta_q[a_{t-q}]\qquad(t=1,2,\cdots,n)$$

我们注意到，若在 (A7.3.4) 中利用 $\hat{\boldsymbol{a}}_*$ 的表达式，则经整理得到

$$S(\boldsymbol{\theta})=\boldsymbol{w}'\boldsymbol{L}_\theta'^{-1}(\boldsymbol{I}_n-\boldsymbol{L}_\theta^{-1}\boldsymbol{F}\boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{L}_\theta'^{-1})\boldsymbol{L}_\theta^{-1}\boldsymbol{w}=\boldsymbol{a}^{0\prime}\boldsymbol{a}^0-\hat{\boldsymbol{a}}_*'\boldsymbol{D}\hat{\boldsymbol{a}}_*$$

这里 $\boldsymbol{a}^0=\boldsymbol{L}_\theta'\boldsymbol{w}$ 表示向量，其元素 a_t^0 可由 $a_t^0=w_t+\theta_1a_{t-1}^0+\cdots+\theta_qa_{t-q}^0, t=1,2,\cdots,n$ 递推计算，且设初值 $\boldsymbol{a}_*$ 等于 0. 因此，上面的第一项 $S_*(\boldsymbol{\theta})=\boldsymbol{a}^{0\prime}\boldsymbol{a}^0=\sum\limits_{t=1}^{n}(a_t^0)^2$ 是如 7.1.2 节所讨论的给出 $\boldsymbol{a}_*=0$ 时的条件平方和函数。

2. 另外，我们得到

$$\boldsymbol{M}_n^{(0,q)}=\boldsymbol{L}_\theta'^{-1}(\boldsymbol{I}_n-\boldsymbol{L}_\theta^{-1}\boldsymbol{F}\boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{L}_\theta'^{-1})\boldsymbol{L}_\theta^{-1}$$

及 $S(\boldsymbol{\theta})=\boldsymbol{w}'\boldsymbol{M}_n^{(0,q)}\boldsymbol{w}$ 。而且通过比较 (A7.3.6) 和 (A7.3.2) 我们有

$$|\boldsymbol{D}|^{-1}=|\boldsymbol{M}_n^{(0,q)}|$$

3. 反向预报 $\hat{\boldsymbol{a}}_*=[\boldsymbol{a}_*]$ 由 $\hat{\boldsymbol{a}}_*=\boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{u}$ 进行计算是最方便的了 (例如，通过解 $\boldsymbol{D}\hat{\boldsymbol{a}}_*=\boldsymbol{F}'\boldsymbol{u}$), 这里， $\boldsymbol{u}=\boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{w}=\boldsymbol{L}_\theta'^{-1}\boldsymbol{a}^0=(u_1,u_2,\cdots,u_n)'$ 。注意， $\boldsymbol{u}$ 的元素 u_t 是通过如下反向递推计算的：

$$u_t=a_t^0+\theta_1u_{t+1}+\cdots+\theta_qu_{t+q}$$

递推从 $t=n$ 到 $t=1$, 并利用了零初始值 $u_{n+1}=\cdots=u_{n+q}=0$, 上式中 a_t^0 记零初始值 $\boldsymbol{a}_*=0$ 条件下 a_t 的估计值，另外，向量 $\boldsymbol{h}=\boldsymbol{F}'\boldsymbol{u}$ 由元素 $h_j=-\sum\limits_{i=1}^{j}\theta_{q-j+i}u_i, j=1,\cdots,q$ 组成。

4. 最后，利用 (A7.3.6), 非条件似然函数由下式精确地给出

$$L(\boldsymbol{\theta},\sigma_a|\boldsymbol{w})=(\sigma_a^2)^{-n/2}|\boldsymbol{D}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_a^2}\sum_{t=1-q}^{n}[a_t]^2\right\}\quad\text{(A7.3.8)}$$

例如，在 MA(q) 的模型中，有 $\boldsymbol{F}'=-(\theta,0,\cdots,0)$ 为 n 维向量，$\boldsymbol{L}_\theta$ 使得 $\boldsymbol{L}_\theta^{-1}$ 的第一列等于 $(1,\theta,\theta^2,\cdots,\theta^{n-1})$, 从而

$$\boldsymbol{D}=1+\boldsymbol{F}'\boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{F}=1+\theta^2+\theta^4+\cdots+\theta^{2n}=\frac{1-\theta^{2(n+1)}}{1-\theta^2}$$

另外，条件值 a_t^0 由 $a_t^0 = w_t + \theta a_{t-1}^9, t = 1, 2, \cdots, n$ 递推计算求得，取初值 $a_0^0 = 0$，而向量 $\boldsymbol{u} = \boldsymbol{L}_\theta'^{-1}\boldsymbol{a}^0$ 由反向递推 $u_t = a_t^0 + \theta u_{t+1}$ 从 $t = n$ 到 $t = 1$ 计算，初值为 $u_{n+1} = 0$ 。于是有

$$\hat{\boldsymbol{a}}_* = [a_0] = -\boldsymbol{D}^{-1}\theta u_1 = -\boldsymbol{D}^{-1}u_0 = -\frac{u_0(1-\theta^2)}{1-\theta^{2(n+1)}}$$

这里 $u_0 = a_0^0 + \theta u_1 = \theta u_1$，于是得到精确似然函数

$$L(\theta, \sigma_a|\boldsymbol{w}) = (\sigma_a^2)^{-n/2}\frac{(1-\theta^2)^{1/2}}{(1-\theta^{2(n+1)})^{1/2}}\exp\left\{-\frac{1}{2\sigma_a^2}\sum_{t=0}^{n}[a_t]^2\right\} \tag{A7.3.9}$$

推广到自回归和混合过程　上面所概括的方法可以很容易地推广到对于混合模型到出非条件似然函数，一般混合模型为

$$\phi(B)\widetilde{w}_t = \theta(B)a_t \tag{A7.3.10}$$

这里，$w_t = \nabla^d z_t$ 定义了一般的 ARIMA 过程。详细的推导已由 Newbold[149] 和 Ljung 和 Box[139] 给出，而另一种求精确似然函数的方法用到带形协方差阵的 Cholesky 分解，这是由 Ansley[11] 给出的。首先假设过程具有零均值，与前文类似， ARMA 模型的关系可以写成矩阵形式:

$$\boldsymbol{L}_\phi\boldsymbol{w} = \boldsymbol{L}_\theta\boldsymbol{a} + \boldsymbol{F}\boldsymbol{e}_*$$

其中 $\boldsymbol{L}_\phi$ 是 $n \times n$ 矩阵与 $\boldsymbol{L}_\theta$ 的形式相同，只需用 ϕ_i 代替 $\theta_i, \boldsymbol{e}_*' = (\boldsymbol{w}_*', \boldsymbol{a}_*') = (w_{1-p}, \cdots, w_0, a_{1-q}, \cdots, a_0)$ 是 $(p+q)$- 维初始值向量，且

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{A}_p & \boldsymbol{B}_q \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

有

$$\boldsymbol{A}_p = \begin{bmatrix} \phi_p & \phi_{p-1} & \cdots & \phi_1 \\ 0 & \phi_p & \cdots & \phi_2 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \phi_p \end{bmatrix} \quad 和 \boldsymbol{B}_q = -\begin{bmatrix} \theta_q & \theta_{q-1} & \cdots & \theta_1 \\ 0 & \theta_q & \cdots & \theta_2 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \theta_q \end{bmatrix}$$

设 $\boldsymbol{\Omega}\sigma_a^2 = E[\boldsymbol{e}_*\boldsymbol{e}_*']$ 记 $\boldsymbol{e}_*$ 的协方差阵。该矩阵表示为

$$\boldsymbol{\Omega}\sigma_a^2 = \begin{bmatrix} \sigma_a^{-2}\boldsymbol{\Gamma}_p & \boldsymbol{C}' \\ \boldsymbol{C} & \boldsymbol{I}_q \end{bmatrix} \sigma_a^2$$

这里, $\boldsymbol{\Gamma}_p = E[\boldsymbol{w}_*\boldsymbol{w}_*']$ 是 $p\times p$ 阵, 其 (i,j) 元素是 γ_{i-j},$\sigma_a^2\boldsymbol{C}=E[\boldsymbol{a}_*\boldsymbol{w}_*']$ 的元素定义为 : $E[a_{i-q}w_{j-p}]=\sigma_a^2\psi_{j-i-p+q}$ 对于 $j-i-p+q>0$, 否则为 0 。 ψ_k 是 MA 无穷算子 $\psi(B)=\phi^{-1}(B)\theta(B)=\sum_{k=0}^{\infty}\psi_k B^k$ 的系数, 其中 $\psi_0=1$, 它们很容易通过 3.4 节中的等式递推地确定出。 $\boldsymbol{\Gamma}_p$ 中的自协方差 γ_k 可以用系数 ϕ_i,θ_i 和 σ_a^2 并通过 (3.4.2) 中的前 $(p+1)$ 个方程直接确定 (见 Ljung 和 Box[139]) 。

类似于 (A7.3.3) 中的结果, 由于 $\boldsymbol{a} = \boldsymbol{L}_\theta^{-1}(\boldsymbol{L}_\phi \boldsymbol{w} - \boldsymbol{F}\boldsymbol{e}_*)$ 和 $\boldsymbol{e}_*$ 是独立的, 故 $\boldsymbol{w}$ 和 $\boldsymbol{e}_*$ 的联合密度是

$$p(\boldsymbol{w},\boldsymbol{e}_*|\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a) = (2\pi\sigma_a^2)^{-(n+p+q)/2}|\boldsymbol{\Omega}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_a^2}S(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{e}_*)\right\}$$

这里

$$S(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{e}_*) = (\boldsymbol{L}_\phi \boldsymbol{w} - \boldsymbol{F}\boldsymbol{e}_*)'\boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}(\boldsymbol{L}_\phi \boldsymbol{w} - \boldsymbol{F}\boldsymbol{e}_*) + \boldsymbol{e}_*'\boldsymbol{\Omega}^{-1}\boldsymbol{e}_*$$

再由广义最小二乘原理我们可以证明

$$S(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{e}_*) = S(\boldsymbol{\phi},\boldsymbol{\theta}) + (\boldsymbol{e}_* - \hat{\boldsymbol{e}}_*)'\boldsymbol{D}(\boldsymbol{e}_* - \hat{\boldsymbol{e}}_*)$$

其中

$$S(\boldsymbol{\phi},\boldsymbol{\theta}) = S(\boldsymbol{\phi},\boldsymbol{\theta},\hat{\boldsymbol{e}}_*) = \hat{\boldsymbol{a}}'\hat{\boldsymbol{a}} + \hat{\boldsymbol{e}}_*'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{e}}_* \qquad (A7.3.11)$$

是非条件平方和函数, 而

$$\hat{\boldsymbol{e}}_* = E[\boldsymbol{e}_*|\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}] = [\boldsymbol{e}_*] = \boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{L}_\phi \boldsymbol{w} \qquad (A7.3.12)$$

是初值 $\boldsymbol{e}_*$ 的条件期望, 其中 $\boldsymbol{D} = \boldsymbol{\Omega}^{-1} + \boldsymbol{F}'\boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{F}, \hat{\boldsymbol{a}} = [\boldsymbol{a}] = \boldsymbol{L}_\theta^{-1}(\boldsymbol{L}_\theta \boldsymbol{w} - \boldsymbol{F}\hat{\boldsymbol{e}}_*)$ 。通过对 $\boldsymbol{w}$ 和 $\boldsymbol{e}_*$ 联合分布的因式分解, 得到

$$p(\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a) = (2\pi\sigma_a^2)^{-n/2}|\boldsymbol{\Omega}|^{-1/2}|\boldsymbol{D}|^{-1/2}\exp\left[-\frac{1}{2\sigma_a^2}S(\boldsymbol{\phi},\boldsymbol{\theta})\right] \qquad (A7.3.13)$$

这是非条件似然函数。由 (A7.3.13) 立刻可得 σ_a^2 的极大似然估计是 $\hat{\sigma}_a^2 = S(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})/n$, 这里， $\hat{\boldsymbol{\phi}}$ 和 $\hat{\boldsymbol{\theta}}$ 表示极大似然估计值。

我们再次注意到， $S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=1}^{n}[a_t]^2 + \hat{\boldsymbol{e}}_*'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{e}}_*$ 和 $\hat{\boldsymbol{a}} = [\boldsymbol{a}]$ 的元素 $[a_1],[a_2],\cdots,[a_n]$ 都是递推计算的，递推关系式是 $\boldsymbol{L}_\theta[\boldsymbol{a}] = \boldsymbol{L}_\phi \boldsymbol{w} - \boldsymbol{F}[\boldsymbol{e}_*]$ 及

$$[a_t] = w_t - \phi_1[w_{t-1}] - \cdots - \phi_p[w_{t-p}] + \theta_1[a_{t-1}] + \cdots + \theta_p[a_{t-q}]$$

对于 $t = 1,2,\cdots,n$, 这里利了初值的反向预报值 $[\boldsymbol{e}_*]$, 以及 $[w_t] = w_t, t = 1,2,\cdots,n$. 另外, 反向预报值 $\hat{\boldsymbol{e}}_* = [\boldsymbol{e}_t]$ 可以由 $\hat{\boldsymbol{e}}_* = \boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{u}$ 计算，其中 $\boldsymbol{u} = \boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{L}_\phi\boldsymbol{w} = \boldsymbol{L}_\theta'^{-1}\boldsymbol{a}^0$ 且 $\boldsymbol{u}$ 的元素 u_t 可通过反向递推计算如下;

$$u_t = a_t^0 + \theta_1 u_{t-1} + \cdots + \theta_q u_{t+q}$$

初始值为 $u_{n+1} = u_{n+q} = \cdots = 0, \boldsymbol{a}^0 = \boldsymbol{L}_\theta^{-1}\boldsymbol{L}_\phi\boldsymbol{w}$ 的元素记为 a_t^0, 它是在零均值 $\boldsymbol{e}_* = 0$ 条件下 a_t 的估计值。此外，向量 $\boldsymbol{h} = \boldsymbol{F}'\boldsymbol{u}$ 由下面的 $p+q$ 个元素组成

$$h_j = \begin{cases} \sum_{i=1}^{j}\phi_{p-j+i}u_i & ,j = 1,\cdots,p \\ -\sum_{i=1}^{j-p}\theta_{q-j+p+i}u_i & ,j = p+1,\cdots,p+q \end{cases}$$

最后, 利用 (A7.1.1) 和 (A7.1.2), 在 $S(\boldsymbol{\phi},\boldsymbol{\theta})$ 中我们记 $\hat{\boldsymbol{e}}_*'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{e}}_* = \hat{\boldsymbol{a}}_*'\hat{\boldsymbol{a}} + (\hat{\boldsymbol{w}}^* - \boldsymbol{C}'\hat{\boldsymbol{a}}_*)'\boldsymbol{K}^{-1}(\hat{\boldsymbol{w}}_*\boldsymbol{C}'\hat{\boldsymbol{a}}_*)$, 于是，我们有

$$S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=1-q}^{n}[a_t]^2 + (\hat{\boldsymbol{w}}_* - \boldsymbol{C}'\hat{\boldsymbol{a}}_*)'\boldsymbol{K}^{-1}(\hat{\boldsymbol{w}}_* - \boldsymbol{C}'\hat{\boldsymbol{a}}_*) \quad \text{(A7.3.14)}$$

其中 $\boldsymbol{K} = \sigma_a^{-2}\boldsymbol{\Gamma}_p - \boldsymbol{C}'\boldsymbol{C}$, 且有 $|\boldsymbol{\Omega}| = |\boldsymbol{K}|$.

因此一般地，若 $\boldsymbol{z}$ 是由任何 ARIMA 过程生成有 $n+d$ 个值的序列，则与其相联系的似然函数由下式给出:

$$L(\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a|\boldsymbol{z}) = (2\pi\sigma_a^2)^{-n/2}|\boldsymbol{M}_n^{(p,q)}|^{1/2}\exp\left[-\frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{2\sigma_a^2}\right] \quad \text{(A7.3.15)}$$

这里

$$S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=1}^{n}[a_t]^2 + \hat{\boldsymbol{e}}_*'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{e}}_*$$

且 $|\boldsymbol{M}_n^{(p,q)}| = |\boldsymbol{\Omega}|^{-1}|\boldsymbol{D}|^{-1} = |\boldsymbol{K}|^{-1}|\boldsymbol{D}|^{-1}$。另外，通过把混合 ARMA 模型考虑为无限阶的滑动平均：$\widetilde{w}_t = (1+\psi_1 B + \psi_2 B^2 + \cdots)a_t$，并参考纯 MA 模型的结果，可知在混合模型的非条件平方和函数中，我们有关系式 $\hat{\boldsymbol{e}}_*'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{e}}_* = \sum_{t=-\infty}^{0}[a_t]^2$。因此，我们也有表达式 $S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=-\infty}^{n}[a_t]^2$，在实际中 $[a_t]$ 的值可从某点 $t = 1-Q$ 开始求和并进行递推计算，而在该点以外的 $[a_t]$ 是可忽略的。

作为特例的自回归过程　在 p 阶纯 AR 模型的特殊情形：$w_t - \sum_{i=1}^{p}\phi_i w_{t-i} = a_t$，前面的结果稍有简化。这时我们有 $\boldsymbol{e}_* = \boldsymbol{w}_*$，$\boldsymbol{\Omega} = \sigma_a^{-2}\boldsymbol{\Gamma}_p$，$\boldsymbol{L}_\theta = \boldsymbol{I}_n$，$\boldsymbol{D} = \sigma_a^2\boldsymbol{\Gamma}_p^{-1} + \boldsymbol{F}'\boldsymbol{F} = \sigma_a^2\boldsymbol{\Gamma}_p^{-1} + \boldsymbol{A}_p'\boldsymbol{A}_p$，以及 $\hat{\boldsymbol{w}}_* = \boldsymbol{D}^{-1}\boldsymbol{F}'\boldsymbol{L}_\phi\boldsymbol{w} = \boldsymbol{D}^{-1}\boldsymbol{A}_p'\boldsymbol{L}_{11}\boldsymbol{w}_p$，这里 $\boldsymbol{w}_p' = (w_1, w_2, \cdots, w_p)$，$\boldsymbol{L}_{11}$ 是 $\boldsymbol{L}_\phi$ 的左上 $p\times p$ 子矩阵。于是可以证明，反向预报值由关系式 $\hat{w}_t = \phi_1\hat{w}_{t+1} + \cdots + \phi_p\hat{w}_{t+p}$，$t = 0, -1, \cdots, -p$ 所确定，其中对 $1 \leqslant t \leqslant n$ 有 $\hat{w}_t = w_t$，因此这与在 7.1.4 节中所讨论的用反向模型方法对 AR 模型特例所得的值一样。这样，我们得到精确平方和为 $S(\boldsymbol{\phi}) = \sum_{t=1}^{n}[a_t]^2 + \sigma_a^2\hat{\boldsymbol{w}}_*'\boldsymbol{\Gamma}_p^{-1}\hat{\boldsymbol{w}}_*$。为了说明起见，考虑 w_t 的一阶自回归过程：

$$w_t - \phi w_{t-1} = a_t \tag{A7.3.16}$$

这里 w_t 可以是实际观测序列的 d 阶差分 $\nabla^d z_t$，并且有一个长度为 $n+d$ 个观测值的序列 $\boldsymbol{z}$。为了计算似然函数 (A7.3.15)，我们要求

$$\begin{aligned}S(\phi) &= \sum_{t=1}^{n}[a_t]^2 + (1-\phi^2)\hat{w}_0^2\\ &= \sum_{t=2}^{n}(w_t - \phi w_{t-1})^2 + (w_1 - \phi\hat{w}_0)^2 + (1-\phi^2)\hat{w}_0^2\end{aligned}$$

其中用到$\boldsymbol{\Gamma}_1 = \gamma_0 = \sigma_a^2(1-\phi^2)^{-1}$。现在由于 $\boldsymbol{D} = \sigma_a^2\boldsymbol{\Gamma}_1^{-1} + \boldsymbol{A}_1'\boldsymbol{A}_1 = \sigma_a^2\gamma_0^{-1} + \phi^2 = 1$, 因此 $\hat{w}_0 = \phi w_1$, 以此代入上式 $S(\phi)$ 的最后两项，则该式简化为

$$S(\phi) = \sum_{t=2}^{n}(w_t - \phi w_{t-1})^2 + (1-\phi^2)w_1^2 \qquad \text{(A7.3.17)}$$

这一结果可由附录 A7.4 中讨论的方法更直接地得到。

以 ARMA(1,1) 作为例子　作为混合模型的例子，我们考虑 ARMA(1,1) 模型：

$$w_t - \phi w_{t-1} = a_t - \theta a_{t-1} \qquad \text{(A7.3.18)}$$

于是我们有 $\boldsymbol{e}_*' = (w_0, a_0), \boldsymbol{A}_1 = \phi, \boldsymbol{B}_1 = -\theta$, 以及

$$\sigma_a^2\boldsymbol{\Omega} = \sigma_a^2\begin{bmatrix}\sigma_a^{-2}\gamma_0 & 1 \\ 1 & 1\end{bmatrix}$$

这里 $\sigma_a^{-2}\gamma_0 = (1+\theta^2-2\phi\theta)/(1-\phi^2)$。因此我们有

$$\begin{aligned}\boldsymbol{D} &= \boldsymbol{\Omega}^{-1} + \boldsymbol{F}'\boldsymbol{L}_\theta'^{-1}\boldsymbol{L}_\theta^{-1}\boldsymbol{F} \\ &= \frac{1}{\sigma_a^{-2}\gamma_0 - 1}\begin{bmatrix}1 & -1 \\ -1 & \sigma_a^{-2}\gamma_0\end{bmatrix} + \frac{1-\theta^{2n}}{1-\theta^2}\begin{bmatrix}\phi^2 & -\phi\theta \\ -\phi\theta & \theta^2\end{bmatrix}\end{aligned}$$

反向预报值由 $\hat{\boldsymbol{e}}_* = \boldsymbol{D}^{-1}\boldsymbol{h}$ 求得，其中 $\boldsymbol{h}' = (h_1, h_2) = (\phi, -\theta)u_1$，而 u_t 是从反向递推 $u_t = a_t^0 + \theta u_{t+1}$ 得到，其中 $u_{n+1} = 0$, 而 $a_t^0 = w_t - \phi w_{t-1}^0 + \theta a_{t-1}^0, t = 1, 2, \cdots, n$ 利用零初值 $w_0^0 = a_0^0 = 0$ 得出，其中 $w_t^0 = w_t, 1 < t < n$。于是求得精确的平方和为

$$S(\phi, \theta) = \sum_{t=0}^{n}[a_t]^2 + \frac{(\hat{w}_0 - \hat{a}_0)^2}{\sigma_a^{-2}\gamma_0 - 1} \qquad \text{(A7.3.19)}$$

其中 $[a_t]=w_t-\phi[w_{t-1}]+\theta[a_{t-1}], t=1,2,\cdots,n$, 且 $\sigma_a^{-2}\gamma_0-1=\boldsymbol{K}=(\phi-\theta)^2/(1-\phi^2)$ 。另外，我们有 $|\boldsymbol{M}_n^{(1,1)}|=\{|\boldsymbol{K}|\,|\boldsymbol{D}|\}^{-1}$, 其中

$$|\boldsymbol{K}|\,|\boldsymbol{D}|=1+\frac{(1-\theta^{2n})(\phi-\theta)^2}{(1-\theta^2)(1-\phi^2)}$$

A 7.4 自回归过程的精确似然函数

现在我们假设，给定序列 $\boldsymbol{w}'=(w_1,w_2,\cdots,w_n)$ 是由下面 p 阶平稳自回归模型生成:

$$w_t-\phi_1w_{t-1}-\phi_2w_{t-2}-\cdots-\phi_pw_{t-p}=a_t$$

这里暂时假设 w 有均值 $\mu=0$, 但是如前所述，有关讨论可推广到 $\mu\neq 0$ 的情形。对 a 假设正态性，从而 w 亦然， w 的联合概率密度函数是

$$p(\boldsymbol{w}|\boldsymbol{\phi},\sigma_a)=(2\pi\sigma_a^2)^{-n/2}|\boldsymbol{M}_n^{(p,0)}|\exp\left[-\frac{\boldsymbol{w}'\boldsymbol{M}_n^{(p,0)}\boldsymbol{w}}{2\sigma_a^2}\right] \quad \text{(A7.4.1)}$$

由于这个一般过程的可逆性，故 $n\times n$ 矩阵 $\boldsymbol{M}_n^{(p,0)}$ 关于它的两个主对角线都是对称的。我们称这样的矩阵为双对称的。现在有

$$p(\boldsymbol{w}|\boldsymbol{\phi},\sigma_a)=p(w_{p+1},w_{p+2},\cdots,w_n|\boldsymbol{w}_p,\boldsymbol{\phi},\sigma_a)p(\boldsymbol{w}_p|\boldsymbol{\phi},\sigma_a)$$

其中 $\boldsymbol{w}_p'=(w_1,w_2,\cdots,w_p)$ 。右边第一个因子可通过使用如下分布求得

$$p(a_{p+1},\cdots,a_n)=(2\pi\sigma_a^2)^{-(n-p)/2}\exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=p+1}^{n}a_t^2\right] \quad \text{(A7.4.2)}$$

对于固定的 $\boldsymbol{w}_p$,$(a_{p+1},\cdots,a_n)$ 和 $(w_{p+1},\cdots,w_n)$ 由以下变换相联系:

$$\begin{aligned}a_{p+1}&=w_{p+1}-\phi_1w_p-\cdots-\phi_pw_1\\&\ \ \vdots\\a_n&=w_n-\phi_1w_{n-1}-\cdots-\phi_pw_{n-p}\end{aligned}$$

该变换具有单位 Jacobi 行列式．因此我们得到

$$
\begin{aligned}
&S(w_{p+1},\cdots,w_n|\boldsymbol{w}_p,\boldsymbol{\phi},\sigma_a)\\
&=(2\pi\sigma_a^2)^{-(n-p)/2}\exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=p+1}^{n}(w_t-\phi_1 w_{t-1}-\cdots-\phi_p w_{t-p})^2\right]
\end{aligned}
$$

及

$$
p(\boldsymbol{w}_p|\boldsymbol{\phi},\sigma_a)=(2\pi\sigma_a^2)^{-p/2}|\boldsymbol{M}_p^{(p,0)}|^{1/2}\exp\left[-\frac{1}{2\sigma_a^2}\boldsymbol{w}_p'\boldsymbol{M}_p^{(p,0)}\boldsymbol{w}_p\right]
$$

因此

$$
p(\boldsymbol{w}|\boldsymbol{\phi},\sigma_a)=(2\pi\sigma_a^2)^{-n/2}|\boldsymbol{M}_p^{(p,0)}|^{1/2}\exp\left[\frac{-S(\boldsymbol{\phi})}{2\sigma_a^2}\right] \tag{A7.4.3}
$$

这里

$$
S(\boldsymbol{\phi})=\sum_{i=1}^{p}\sum_{j=1}^{p}m_{ij}^{(p)}w_i w_j+\sum_{t=p+1}^{n}(w_t-\phi_1 w_{t-1}-\cdots-\phi_p w_{t-p})^2 \tag{A7.4.4}
$$

并有

$$
\boldsymbol{M}_p^{(p,0)}=\{m_{ij}^{(p)}\}=\{\gamma_{|i-j|}\}^{-1}\sigma_a^2=\begin{bmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{p-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2}\\ \vdots & \vdots & & \vdots\\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0\end{bmatrix}^{-1}\sigma_a^2 \tag{A7.4.5}
$$

这里 $\gamma_0,\gamma_1,\cdots,\gamma_{p-1}$ 为过程的理论自协方差，且 $|\boldsymbol{M}_p^{(p,0)}|=|\boldsymbol{M}_n^{(p,0)}|$。

现在令 $n=p+1$，于是有

$$
\begin{aligned}
\boldsymbol{w}_{p+1}'\boldsymbol{M}_{p+1}^{(p,0)}\boldsymbol{w}_{p+1}=&\sum_{i=1}^{p}\sum_{j=1}^{p}m_{ij}^{(p)}w_i w_j\\
&+(w_{p+1}-\phi_1 w_p-\phi_2 w_{p-1}-\cdots-\phi_p w_1)^2
\end{aligned}
$$

那么

$$
\boldsymbol{M}_{p+1}^{(p)}=\begin{bmatrix} & & & \vdots & 0 \\ & & & \vdots & 0 \\ & \boldsymbol{M}_p^{(p)} & & \vdots & \vdots \\ & & & \vdots & \vdots \\ \cdots & \cdots & \cdots & \vdots & \cdots \\ 0 & 0 & \cdots & \vdots & 0 \end{bmatrix}+\begin{bmatrix} \phi_p^2 & \phi_p\phi_{p-1} & \cdots & \vdots & -\phi_p \\ \phi_p\phi_{p-1} & \phi_{p-1}^2 & \cdots & \vdots & -\phi_{p-1} \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & -\phi_1 \\ \cdots & \cdots & \cdots & \vdots & \cdots \\ -\phi_p & -\phi_{p-1} & \cdots & \vdots & 1 \end{bmatrix}
$$

通过考虑到 $\boldsymbol{M}_p^{(p)}$ 和 $\boldsymbol{M}_{p+1}^{(p)}$ 都是双对称的，于是可将 $\boldsymbol{M}_p^{(p)}=\boldsymbol{M}_p^{(p,0)}$ 的元素推出。因此，例如

$$
\boldsymbol{M}_2^{(1)}=\begin{bmatrix} m_{11}^{(1)}+\phi_1^2 & -\phi_1 \\ -\phi_1 & 1 \end{bmatrix}=\begin{bmatrix} 1 & -\phi_1 \\ -\phi_1 & m_{11}^{(1)}+\phi_1^2 \end{bmatrix}
$$

令两矩阵中元素相等，我们有

$$
\boldsymbol{M}_1^{(1)}=m_{11}^{(1)}=1-\phi_1^2
$$

依此处理，我们对 1 阶和 2 阶过程求得

$$
\begin{aligned}
\boldsymbol{M}_1^{(1)}&=1-\phi_1^2 \quad , \quad |\boldsymbol{M}_1|^{(1)}=1-\phi_1^2 \\
\boldsymbol{M}_2^{(2)}&=\begin{bmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{bmatrix} \\
|\boldsymbol{M}_2^{(2)}|&=(1+\phi_2)^2[(1-\phi_2)^2-\phi_1^2]
\end{aligned}
$$

例如当 $p=1$ 时有

$$
\begin{aligned}
p(\boldsymbol{w}|\phi,\sigma_a)=&(2\pi\sigma_a^2)^{-n/2}(1-\phi^2)^{1/2} \\
&\times\exp\left\{-\frac{1}{2\sigma_a^2}\left[(1-\phi^2)w_1^2+\sum_{t=2}^{n}(w_t-\phi w_{t-1})^2\right]\right\}
\end{aligned}
$$

这检验了 (A7.3.17) 的结果。上述生成的过程必然会得到矩阵 $\boldsymbol{M}_p^{(p)}$，其元素为 ϕ 的二次型。

因此由 (A7.4.4) 显然可见，$S(\boldsymbol{\phi})=\boldsymbol{w}'\boldsymbol{M}_n^{(p)}\boldsymbol{w}$ 不仅是 w 的二次型，而且也是参数 $\boldsymbol{\phi}$ 的二次型。记 $\boldsymbol{\phi}_u'=(1,\phi_1,\phi_2,\cdots,\phi_p)$，设矩阵 $\boldsymbol{D}$ 是某个 $(p+1)\times(p+1)$ 阵，其元素是 w 的二次函数，则下式显然成立

$$\boldsymbol{w}'\boldsymbol{M}_n^{(p)}\boldsymbol{w}=\boldsymbol{\phi}_u'\boldsymbol{D}\boldsymbol{\phi}_u$$

现在记

$$\boldsymbol{D}=\begin{bmatrix} D_{11} & -D_{12} & -D_{13} & \cdots & -D_{1,p+1} \\ -D_{12} & D_{22} & D_{23} & \cdots & D_{2,p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ -D_{1,p+1} & D_{2,p+1} & D_{3,p+1} & \cdots & D_{p+1,p+1} \end{bmatrix} \tag{A7.4.6}$$

对 (A7.4.4) 的考察表明，D_{ij} 是滞后乘积及平方的对称和，定义为

$$D_{ij}=D_{ji}=w_iw_j+w_{i+1}w_{j+1}+\cdots+w_{n+1-j}w_{n+1-i} \tag{A7.4.7}$$

其中，和式 D_{ij} 包含 $n-(i-1)-(j-1)$ 项。

最后，我们可以写出精确的概率密度，也即精确的似然函数，它们是

$$p(\boldsymbol{w}|\boldsymbol{\phi},\sigma_a)=L(\boldsymbol{\phi},\sigma_a|\boldsymbol{w})=(2\pi\sigma_a^2)^{-n/2}|\boldsymbol{M}_p^{(p)}|^{1/2}\exp\left[\frac{-S(\boldsymbol{\phi})}{2\sigma_a^2}\right] \tag{A7.4.8}$$

其中

$$S(\boldsymbol{\phi})=\boldsymbol{w}_p'\boldsymbol{M}_p^{(p)}\boldsymbol{w}_p+\sum_{t=p+1}^{n}(w_t-\phi_1w_{t-1}-\cdots-\phi_pw_{t-p})^2=\boldsymbol{\phi}_u'\boldsymbol{D}\boldsymbol{\phi}_u \tag{A7.4.9}$$

对数似然函数为

$$l(\boldsymbol{\phi},\sigma_a|\boldsymbol{w})=-\frac{n}{2}\ln(\sigma_a^2)+\frac{1}{2}\ln|\boldsymbol{M}_p^{(p)}|-\frac{S(\boldsymbol{\phi})}{2\sigma_a^2} \tag{A7.4.10}$$

极大似然估计　在 (A7.4.10) 中对每个 ϕ 求导，我们得到

$$\frac{\partial l}{\partial \sigma_a}=-\frac{n}{\sigma_a}+\frac{S(\boldsymbol{\phi})}{\sigma_a^3} \tag{A7.4.11}$$

$$\frac{\partial l}{\partial \phi_j}=M_j+\sigma_a^{-2}(D_{1,j+1}-\phi_1 D_{2,j+1}-\cdots-\phi_p D_{p+1,j+1}),\quad (j=1,2,\cdots,p) \tag{A7.4.12}$$

其中

$$M_j=\frac{\partial(\frac{1}{2}\ln|\boldsymbol{M}_p^{(p)}|)}{\partial \phi_j}$$

因此，通过令这些表达式为 0 并对得到的方程求解，便可得到极大似然估计。

由 (A7.4.1) 我们立刻有

$$\hat{\sigma}_a^2=\frac{S(\hat{\boldsymbol{\phi}})}{n} \tag{A7.4.13}$$

$\boldsymbol{\phi}$ 的估计　在处理方程 (A7.4.12) 时会有一定的困难发生，因为通常 $\boldsymbol{M}_j(j=1,2,\cdots,p)$ 是 ϕ 的复杂函数。在这里我们考虑其它的四种简化方法。

1. 最小二乘估计。

虽然 $S(\boldsymbol{\phi})$ 的期望值与 n 成比例，$|\boldsymbol{M}_p^{(p)}|$ 的值与 n 独立，且对于中等和大样本情形，(A7.4.8) 被 $S(\boldsymbol{\phi})$ 和 $|\boldsymbol{M}_p^{(p)}|$ 中的项所控制，而 $|\boldsymbol{M}_p^{(p)}|$ 中的项相比起来更小。

如果我们忽略该项的影响，则有

$$l(\boldsymbol{\phi},\sigma_a|\boldsymbol{w})\simeq-\frac{n}{2}\ln(\sigma_a^2)-\frac{S(\boldsymbol{\phi})}{2\sigma_a^2} \tag{A7.4.14}$$

由极大化 (A7.4.14) 得到 $\boldsymbol{\phi}$ 的估计值 $\hat{\boldsymbol{\phi}}$ 就是极小化 $S(\boldsymbol{\phi})$ 所得的最小二乘估计。现在，由 (A7.4.9) 有 $S(\boldsymbol{\phi})=\boldsymbol{\phi}_u'\boldsymbol{D}\boldsymbol{\phi}_u$，这里 $\boldsymbol{D}$ 是平

方和乘积对称和的 $(p+1)\times(p+1)$ 矩阵。因此，通过微分，最小化值为

$$
\begin{array}{lllll}
D_{12} & =\hat{\phi}_1 D_{22} & +\hat{\phi}_2 D_{23} & +\cdots+\hat{\phi}_p D_{2,p+1} \\
D_{13} & =\hat{\phi}_1 D_{23} & +\hat{\phi}_2 D_{33} & +\cdots+\hat{\phi}_p D_{3,p+1} \\
\vdots & \vdots & \vdots & \vdots \\
D_{1,p+1} & =\hat{\phi}_1 D_{2,p+1} & +\hat{\phi}_2 D_{3,p+1} & +\cdots+\hat{\phi}_p D_{p+1,p+1}
\end{array}
\qquad (A7.4.15)
$$

用清楚的矩阵记号可以写成

$$
\boldsymbol{d}=\boldsymbol{D}_p\hat{\boldsymbol{\phi}}
$$

于是有

$$
\hat{\boldsymbol{\phi}}=\boldsymbol{D}_p^{-1}\boldsymbol{d}
$$

这些最小二乘估计也使后验密度 (7.4.15) 极大化。

2. 近似极大似然估计

现在我们来回忆一个以前的结果(3.2.3)，它可写为

$$
\gamma_j-\phi_1\gamma_{j-1}-\phi_2\gamma_{j-2}-\cdots-\phi_p\gamma_{j-p}=0 \quad (j>0) \qquad (A7.4.16)
$$

另外，通过在 (A7.4.12) 中取期望，并利用 $E[\partial l/\partial\phi_j]=0$ 这一事实，我们得出

$$
\begin{aligned}
M_j\sigma_a^2+(n-j)\gamma_j-(n-j-1)\phi_j\gamma_{j-1}-(n-j-2)\phi_2\gamma_{j-2} \\
-\cdots-(n-j-p)\phi_p\gamma_{j-p}=0
\end{aligned}
\qquad (A7.4.17)
$$

(A7.4.16) 乘以 n, 并从 (A7.4.17) 将其减去，我们得到

$$
M_j\sigma_a^2=j\gamma_j-(j+1)\phi_1\gamma_{j-1}-\cdots-(j+p)\phi_p\gamma_{j-p}
$$

因此，通过用 $D_{i+1,j+1}/(n-j-i)$ 作为 $\gamma_{|j-i|}$ 的估计值，那么，$M_j\sigma_a^2$ 的一个很自然的估计值就是

$$
j\frac{D_{1,j+1}}{n-j}-(j+1)\phi_1\frac{D_{2,j+1}}{n-j-1}-\cdots-(j+p)\phi_p\frac{D_{p+1,j+1}}{n-j-p}
$$

在 (A7.4.12) 中代入这个估计值，得出

$$\frac{\partial l}{\partial \phi_j} \simeq n\sigma_a^{-2}\left(\frac{D_{1,j+1}}{n-j} - \phi_1\frac{D_{2,j+1}}{n-j-1} - \cdots - \phi_p\frac{D_{p+1,j+1}}{n-j-p}\right)$$

$$(j = 1, 2, \cdots, p) \qquad \text{(A7.4.18)}$$

于是推出一组形如 (A7.4.15) 的线性方程，只不过现在用

$$D_{ij}^* = \frac{nD_{ij}}{n-(i-1)-(j-1)}$$

来代替 D_{ij} 。

3. 条件最小二乘估计

对于中等和相对较大的 n, 我们也可考虑条件平方和函数，这是用在 7.1.3 节中的方法所得到的。这样，对于表达式 (A7.4.2) 指数中的平方和有

$$S_*(\boldsymbol{\phi}) = \sum_{t=p+1}^{n}(w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p})^2$$

这就是在给定 $\boldsymbol{w}_p' = (w_1, w_2, \cdots, w_p)$ 时与 $w_{p+1}, \cdots, w_n$ 的条件分布相联系的平方和。条件最小二乘估计是通过极小化 $S_*(\boldsymbol{\phi})$ 而得到，这是一个与线性模型 $w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} + a_t, t = p+1, \cdots, n$ 相联系的线性最小二乘回归问题。这就得到人们所熟悉的最小二乘估计 $\hat{\boldsymbol{\phi}} = \widetilde{\boldsymbol{D}}_p^{-1}\widetilde{\boldsymbol{d}}$, 与在 (A7.2.5) 中一样，这里 $\widetilde{\boldsymbol{D}}_p$ 的第 (i,j) 元素是 $\widetilde{D}_{ij} = \sum_{t=p+1}^{n} w_{t-i}w_{t-j}$,$\widetilde{\boldsymbol{d}}$ 的第 i 元素是 $\widetilde{d}_i = \sum_{t=p+1}^{n} w_{t-i}w_t$ 。

4. Yule-Walker 估计

最后，如果 n 是中等或较大的，作为一种近似，我们可以用适当的 n 倍自协方差估计值来代替 (A7.4.15) 中对称的平方与乘

积和。例如 D_{ij}(这里 $|i-j|=k$) 将用 $nc_k=\sum_{t=1}^{n-k}\widetilde{w}_t\widetilde{w}_{t+k}$ 。用 nc_0 去除所有得到的方程，我们便得到以自相关函数估计值 $r_k=c_k/c_0$ 表示的如下关系

$$\begin{aligned}
r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 + \cdots + \hat{\phi}_p r_{p-1}\\
r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 + \cdots + \hat{\phi}_p r_{p-2}\\
&\vdots\\
r_p &= \hat{\phi}_1 r_{p-1} + \hat{\phi}_2 r_{p-2} + \cdots + \hat{\phi}_p
\end{aligned}$$

这就是众所周知的 Yule-Walker 方程。

在矩阵 (7.3.1) 中它们可以被写为 $\boldsymbol{r}=\boldsymbol{R}\hat{\boldsymbol{\phi}}$, 于是

$$\hat{\boldsymbol{\phi}}=\boldsymbol{R}^{-1}\boldsymbol{r} \tag{A7.4.19}$$

这相应于方程 (3.2.7), 只是用 $\boldsymbol{r}$ 来代替 $\boldsymbol{\rho}_p$, 用 $\boldsymbol{R}$ 代替 $\boldsymbol{P}_p$ 。

为了说明这四个估计的差别，我们以 $p=1$ 为例加以说明。于是相应于 (A7.4.12) 有 $M_1\sigma_a^2=-\gamma_1$,ϕ 的极大似然估计是下式的解

$$-\gamma_1+\sum_{t=2}^{n}w_tw_{t-1}-\phi\sum_{t=2}^{n-1}w_t^2=0$$

注意， $\gamma_1=\sigma_a^2\phi/(1-\phi^2)$, 并且用如 (A7.4.11) 得到的极大似然解 $\hat{\sigma}_a^2=S(\phi)/n$, 其中 $S(\phi)=D_{11}-2\phi D_{12}+\phi^2D_{22}$ 来替换上面似然方程中 γ_1 表达式中的 σ_a^2, 我们就可得到 ϕ 的一个三次方程，其解便是 ϕ 的极大似然估计。

近似方法 (1) 相当于完全略去 γ_1 项，得到

$$\hat{\phi}\doteq\frac{\sum\limits_{t=2}^{n}w_tw_{t-1}}{\sum\limits_{t=2}^{n-1}w_t^2}=\frac{D_{12}}{D_{22}}$$

(2) 相当于用估计值 $\sum_{t=2}^{n} w_t w_{t-1}/(n-1)$ 代替 γ_1, 得到

$$\hat{\phi} = \frac{\sum_{t=2}^{n} w_t w_{t-1}/(n-1)}{\sum_{t=2}^{n-1} w_t^2/(n-2)} = \frac{n-2}{n-1}\frac{D_{12}}{D_{22}}$$

(3) 相当于用 w_t 在 $w_{t-1}, t = 2, 3, \cdots, n$ 上作回归所得的标准线性模型的最小二乘估计，于是有

$$\hat{\phi} = \frac{\sum_{t=2}^{n} w_t w_{t-1}}{\sum_{t=2}^{n} w_{t-1}^2} = \frac{D_{12}}{D_{22} + w_1^2}$$

(4) 是用标准自协方差估计 (2.1.10) 替换分子和分母，得到

$$\hat{\phi} = \frac{\sum_{t=2}^{n} w_t w_{t-1}}{\sum_{t=2}^{n-1} w_t^2} = \frac{c_1}{c_0} = r_1 = \frac{D_{12}}{D_{22}}$$

通常，与该例一样，对中等及大样本，各种近似方法给出的估计值之间差别很小。我们一般都用 (1) 给出的最小二乘估计。当然这些估计值可直接由 (A7.4.15) 计算。当有计算软件包可以利用时，单独处理自回归过程几乎是无意义的，我们发现，即使在拟合自回归过程时，使用 7.2.1 节中所述计算任意 ARMA 过程的一般迭代算法也是最简便的。

σ_a^2 的估计　对 (A7.4.9) 和 (A7.4.13) 运用近似 (4), 得

$$\hat{\sigma}_a^2 = \frac{S(\hat{\boldsymbol{\phi}})}{n} = c_0(1 \vdots \hat{\boldsymbol{\phi}}') \begin{bmatrix} 1 & \vdots & -\boldsymbol{r}' \\ \cdots & \vdots & \cdots \\ -\boldsymbol{r} & \vdots & \boldsymbol{R} \end{bmatrix} \begin{pmatrix} 1 \\ \cdots \\ \hat{\boldsymbol{\phi}} \end{pmatrix}$$

将右边乘出，并注意 $\boldsymbol{r}-\boldsymbol{R}\hat{\boldsymbol{\phi}}=0$, 我们得到

$$\hat{\sigma}_a^2=c_0(1-\boldsymbol{r}'\hat{\boldsymbol{\phi}})=c_0(1-\boldsymbol{r}'\boldsymbol{R}^{-1}\boldsymbol{r})=c_0(1-\hat{\boldsymbol{\phi}}'\boldsymbol{R}\hat{\boldsymbol{\phi}}) \quad \text{(A7.4.20a)}$$

已经证明，σ_a^2 可由理论相关函数用相似的形式写出

$$\sigma_a^2=\gamma_0(1-\boldsymbol{\rho}'\boldsymbol{\phi})=\gamma_0(1-\boldsymbol{\rho}'\boldsymbol{P}_p^{-1}\boldsymbol{\rho})=\gamma_0(1-\boldsymbol{\phi}'\boldsymbol{P}_p\boldsymbol{\phi}) \quad \text{(A7.4.20b)}$$

这与 (3.2.8) 的结果一致。

信息阵　对 (A7.4.11) 和 (A7.4.18) 作两次微分，我们得到

$$-\frac{\partial^2 l}{\partial\sigma_a^2}=-\frac{n}{\sigma_a^2}+\frac{3S(\boldsymbol{\phi})}{\sigma_a^4} \quad \text{(A7.4.21a)}$$

$$\frac{\partial^2 l}{\partial\sigma_a\partial\phi_j}\simeq-2\sigma_a^{-1}\frac{\partial l}{\partial\phi_j} \quad \text{(A7.4.21b)}$$

$$\frac{\partial^2 l}{\partial\phi_i\partial\phi_j}\simeq\frac{n}{\sigma_a^2}\frac{D_{i+1,j+1}}{n-i-j} \quad \text{(A7.4.21c)}$$

现在，由于

$$E\left[\frac{\partial l}{\partial\phi_j}\right]=0$$

对于中等或大样本来说，得到

$$E\left[-\frac{\partial^2 l}{\partial\sigma_a\partial\phi_j}\right]\simeq 0$$

及

$$|\boldsymbol{I}(\boldsymbol{\phi},\sigma_a)|\simeq|\boldsymbol{I}(\boldsymbol{\phi})|I(\sigma_a)$$

其中

$$I(\sigma_a)=E\left[-\frac{\partial^2 l}{\partial\sigma_a^2}\right]=\frac{2n}{\sigma_a^2}$$

现在，利用 (A7.4.21), 我们有

$$\boldsymbol{I}(\boldsymbol{\phi})=-E\left[\frac{\partial^2 l}{\partial\phi_i\partial\phi_j}\right]\simeq\frac{n}{\sigma_a^2}\boldsymbol{\Gamma}_p=\frac{n\gamma_0}{\sigma_a^2}\boldsymbol{P}_p=n(\boldsymbol{M}_p^{(p)})^{-1} \quad \text{(A7.4.22)}$$

因此

$$|\boldsymbol{I}(\boldsymbol{\phi},\sigma_a)| \simeq \frac{2n^{p+1}}{\sigma_a^2}|\boldsymbol{M}_p^{(p,0)}|^{-1}$$

自回归参数估计值的方差和协方差　在 [196] 所充分讨论的情形中，信息阵的逆给出了极大似然 (ML) 估计的渐近方差 - 协方差阵。而且，如果对数似然函数近似为二次型，且最大值不接近边界，那么，即使样本数目是中等大小的，这个矩阵的元素通常也会给出估计值的方差和协方差的充分近似。

因此，由 (A7.4.33) 和 (A7.4.20) 给出了

$$\begin{aligned}\boldsymbol{V}(\boldsymbol{\phi}) = \boldsymbol{I}^{-1}(\hat{\boldsymbol{\phi}}) \simeq n^{-1}\boldsymbol{M}_p^{(p)} &= n^{-1}\sigma_a^2\boldsymbol{\Gamma}_p^{-1}\\ &= n^{-1}(1-\boldsymbol{\rho}'\boldsymbol{P}_p^{-1}\boldsymbol{\rho})\boldsymbol{P}_p^{-1}\\ &= n^{-1}(1-\boldsymbol{\phi}'\boldsymbol{P}_p\boldsymbol{\phi})\boldsymbol{P}_p^{-1}\\ &= n^{-1}(1-\boldsymbol{\rho}'\boldsymbol{\phi})\boldsymbol{P}_p^{-1}\end{aligned} \tag{A7.4.23}$$

特别地，对于一阶和二阶自回归过程有

$$\begin{aligned}V(\hat{\phi}) &\simeq n^{-1}(1-\phi^2)\\ \boldsymbol{V}(\hat{\phi}_1,\hat{\phi}_2) &\simeq n^{-1}\begin{bmatrix}1-\phi_2^2 & -\phi_1(1+\phi_2)\\ -\phi_1(1+\phi_2) & 1-\phi_2^2\end{bmatrix}\end{aligned} \tag{A7.4.24}$$

方差和协方差的估计值可以在 (A7.4.24) 中用估计值来代替参数而得出。例如，我们可以在 (A7.4.24) 中用 r 代替 ρ, 用 $\hat{\boldsymbol{\phi}}$ 代替 $\boldsymbol{\phi}$ 而得到

$$\hat{\boldsymbol{V}}(\hat{\boldsymbol{\phi}}) \simeq n^{-1}(1-\boldsymbol{r}'\hat{\boldsymbol{\phi}})\boldsymbol{R}^{-1} \tag{A7.4.25}$$

A 7.5　参数估计误差对预报概率限影响的例子

5.2.4 节中对预报值给出了方差和概率限，这是建立在 ARIMA 模型的参数 $(\boldsymbol{\phi},\boldsymbol{\theta})$ 为确切知道的基础之上的，而在实际中，必须

用它们的估计值 $(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})$ 来代替．为了更进一步理解估计误差对预报误差方差的影响，我们考虑非平稳 IMA(0,1,1) 过程和平稳一阶自回归过程的特殊情形．对这两个过程及建立在中等长度序列基础上的参数估计表明，这种估计误差的影响是较小的。

IMA(0,1,1) 过程　将模型 $\nabla z_t = a_t - \theta a_{t-1}$, 对 $t+l, t+l-1,\cdots,t+1$ 写出，并进行求和，于是我们得到

$$z_{t+l} - z_t = a_{t+l} + (1-\theta)(a_{t+l-1} + \cdots + a_{t+1}) - \theta a_t$$

当参数 θ 确切知道时，用 $\hat{z}_t(l|\theta)$ 记向前 l 期预报．在时刻 t 对 $l=1,2,\cdots$ 取条件期望，我们得到

$$\begin{aligned}\hat{z}(1|\theta) &= z_t - \theta z_t \\ \hat{z}(l|\theta) &= \hat{z}_t(1|\theta) \qquad (l > 2)\end{aligned}$$

因此，向前 l 期预报误差为

$$\begin{aligned}e_t(l|\theta) &= z_{t+l} - \hat{z}_t(l|\theta) \\ &= a_{t+l} + (1-\theta)(a_{t+l-1} + \cdots + a_{t+1})\end{aligned}$$

而提前 l 期预报误差的方差是

$$V(l) = E_t[e_t^2(l|\theta)] = \sigma_a^2[1 + (l-1)\lambda^2] \qquad \text{(A7.5.1)}$$

这里 $\lambda = 1-\theta$ 。

因此，若对于 $w_t = \nabla z_t$ 的 n 个值构成的时间序列，得到了估计值 $\hat{\theta}$, 并由它来代替 θ, 则有

$$\begin{aligned}\hat{z}_t(1|\hat{\theta}) &= z_t - \hat{\theta}\hat{a}_t \\ \hat{z}_t(l|\hat{\theta}) &= \hat{z}_t(1|\hat{\theta}) \qquad (l > 2)\end{aligned}$$

其中 $\hat{a}_t = z_t - \hat{z}_{t-1}(1|\hat{\theta})$ ．因此由 $\hat{\theta}$ 所得的提前 l 期预报误差为

$$\begin{aligned}e_t(1|\hat{\theta}) &= z_{t+l} - \hat{z}_t(l|\hat{\theta}) \\ &= z_{t+l} - z_t + \hat{\theta}\hat{a}_t \qquad \text{(A7.5.2)} \\ &= e_t(l|\theta) - (\theta a_t - \hat{\theta}\hat{a}_t)\end{aligned}$$

因为 $\nabla z_t = (1-\theta B)a_t = (1-\hat{\theta}B)\hat{a}_t$, 可得出

$$\hat{a}_t = \left(\frac{1-\theta B}{1-\hat{\theta}B}\right) a_t$$

并且从 (A7.5.2) 中消去 $\hat{a}_t$, 我们写成

$$e_t(l|\hat{\theta}) = e_t(l|\theta) - \frac{\theta-\hat{\theta}}{1-\hat{\theta}B} a_t$$

那么

$$\begin{aligned}
\frac{\theta-\hat{\theta}}{1-\hat{\theta}B} a_t &= \frac{\theta-\hat{\theta}}{1-\theta B}\left[1+\frac{(\theta-\hat{\theta})B}{1-\theta B}\right]^{-1} a_t \\
&\simeq \frac{\theta-\hat{\theta}}{1-\theta B}\left[1+\frac{(\theta-\hat{\theta})B}{1-\theta B}\right] a_t \qquad \text{(A7.5.3)} \\
&= (\theta-\hat{\theta})(a_t+\theta a_{t-1}+\theta^2 a_{t-2}+\cdots) \\
&\quad -(\theta-\hat{\theta})^2(a_{t-1}+2\theta a_{t-2}+3\theta^2 a_{t-3}+\cdots)
\end{aligned}$$

假设预报值和估计值 $\hat{\theta}$ 是建立在基本上不重迭数据的基础上，则 $\hat{\theta}$ 和 $a_t, a_{t-1}, \cdots$ 是独立的。且对于中等大小样本来说，$\hat{\theta}$ 是以 θ 为中心、方差为 $(1-\theta^2)/n$ 的近似正态分布。在这些假设之下可以证明 (A7.5.3) 中表达式的方差为

$$\frac{\sigma_a^2}{n}\left(1+\frac{3}{n}\frac{1+\theta^2}{1-\theta^2}\right)$$

因此，若 $|\theta|$ 不接近于 1, 则

$$\text{var}[e_t(l|\hat{\theta})] \simeq \sigma_a^2[1+(l-1)\lambda^2] + \frac{\sigma_a^2}{n} \qquad \text{(A7.5.4)}$$

显然，当精确的预报误差下降到 σ_a^2 时，对于 $l=1$ 方差中的比例变化是最大的。在这种情形，对基于中等长度序列的参数估计，概率限将按倍数 $(n+1)/n$ 增加。

一阶自回归过程　在时刻 $t+1$ 写出模型 $\tilde{z}_t=\phi\tilde{z}_{t-1}+a_t$ 并关于时刻 t 取条件期望，若参数的真值已给出，则提前 l 期的预报值是

$$\hat{\tilde{z}}_t(l|\phi)=\phi\hat{\tilde{z}}_t(l-1|\phi)=\phi^l\tilde{z}_t$$

类似地，有

$$\hat{\tilde{z}}_t(l|\hat{\phi})=\hat{\phi}\hat{\tilde{z}}_t(l-1|\hat{\phi})=\hat{\phi}^l\tilde{z}_t$$

因此有

$$e_t(l|\hat{\phi})=e_t(l|\phi)+(\phi^l-\hat{\phi}^l)\tilde{z}_t$$

所以，

$$E_t[e_t^2(l|\hat{\phi})]=E_t[e_t^2(l|\phi)]+\tilde{z}_t^2E_t[(\phi^l-\hat{\phi}^l)^2]$$

于是一般说来有

$$\mathrm{var}[e_t(l|\hat{\phi})]\simeq\sigma_a^2\frac{(1-\phi^{2l})}{(1-\phi^2)}+\sigma_a^2\frac{E[(\phi^l-\hat{\phi}^l)^2]}{1-\phi^2} \tag{A7.5.5}$$

利用 (5.4.16), 当 $l=1$ 时，

$$\begin{aligned}\mathrm{var}[e_t(1|\hat{\phi})]&\simeq\sigma_a^2+\frac{\sigma_a^2}{1-\phi^2}\,\frac{1-\phi^2}{n}\\&=\sigma_a^2\left(1+\frac{1}{n}\right)\end{aligned} \tag{A7.5.6}$$

对于 $l>1$, 我们有

$$\phi^l-\hat{\phi}^l=\phi^l-\{\phi-(\phi-\hat{\phi})\}^l=\phi^l-\phi^l\left(1-\frac{\phi-\hat{\phi}}{\phi}\right)^l\simeq l\phi^{l-1}(\phi-\hat{\phi})$$

因此，一般有，

$$\mathrm{var}[e_t(l|\hat{\phi})]\simeq\mathrm{var}[e_t(l|\phi)]+\frac{l^2\phi^{2(l-1)}}{n}\sigma_a^2$$

且偏差的量级仍为 n^{-1} 。

对于一般的 AR(p) 模型，Yamamoto[210] 给出了参数估计误差对预报方差影响的有关近似结果。特别地，对于 p 阶 AR 模型的一步预报，立刻可以得到下面的近似式：

$$\operatorname{var}[e_t(1|\hat{\boldsymbol{\phi}})] \simeq \sigma_a^2\left(1+\frac{p}{n}\right)$$

A 7.6 关于滑动平均参数估计的特别注记

如果最小二乘迭代允许伸延到可逆域以外，则可以找到相应平方和比真正最小值更小的参数。然而，这并不能给出合理的估计，而且是毫无意义的。为了说明起见，假设一个序列由一阶滑动平均模型 $w_t=(1-\theta B)a_t$ 生成，且 $-1<\theta<1$ 。那么，该序列等价于由相应的反向过程 $w_t=(1-\theta F)e_t$ 产生，其中 $\sigma_e^2=\sigma_a^2$ 。现在后一过程也可记为 $w_t=(1-\theta^{-1}B)\alpha_t$，而此处的 θ^{-1} 是在可逆域以外。然而，在该表达式中 $\sigma_\alpha^2=\sigma_a^2\theta^2$，该量本身就是 θ 的函数，因此，θ^{-1} 的有效估计将不能由 $\sum\alpha_t^2=\theta^2\sum a_t^2$ 的极小化给出。事实上，上式在 $\theta^{-1}=\infty$ 处达到极小。

上述困难是可以避免的：

1. 通过用在模型识别阶段得到在可逆域内的初始估计作为初估计。

2. 通过检验收敛已经发生后所得的滑动平均估计值，这些估计值应在可逆域内。

另外，还可以在编写最小二乘程序时使估计值限制在可逆域内。

8
模型的诊断检验

经过模型识别和参数估计后，随之要对模型进行诊断检验．检验模型的一种有效方法是“过拟合”，也就是对于我们信其为真的一类模型，我们用稍许广泛一些的模型去估计其参数．这种方法假设我们能够猜出模型大概在哪些方面不恰当，因此，必须采用一些方法对拟合模型的残差做不很特殊的检验．使得我们可以从数据本身出发，提出对于模型修改的建议．我们将要讲述的就是这样的检验方法，所使用的工具是 (1) 残差的自相关函数，(2) 残差的累计周期图．

8.1 随机模型的检验

8.1.1 基本原理

假设使用一个特定的时间序列，运用第 6 章和第 7 章的方法，我们已经进行了模型识别和参数估计．剩下的问题就是要确定模型是否恰当．如果存在拟合严重不当的证据，那么就需要弄清楚在下一环节中应该如何去修改模型．我们着手做的这些工作只能用“检验拟合优度”一词来做不全面的描述．我们需要揭示出模型到底怎么不恰当，并提出如何进行适当的修改．为了说明问题，将要参考时间序列以外的一些熟知的方法，例如，由 Anscombe 和 Tukey [9],[10] 所述有关方差分析的残差检验，以及由 Daniel[76]

引入的用正态图和其它方法的因子试验鉴别，所有这些我们都称之为诊断检验。

从来没有一种模型能绝对真实地描述客观现实。所以在给出充足数据的前提下，统计检验就是对于当下使用充分适合的模型提出质疑。从另一方面说，检验也可能不能指出模型对假设有严重偏离，这是因为这些检验对所发生偏离的类型是不灵敏的。因此，最好的方针就是设计出尽可能灵敏的统计方法，但是，对于一些拟合精度明显不足的模型，出于某些充分的理由，我们仍然准备使用它，那么，就要尽可能地弄清楚客观事实，然后加以判断。

显然，诊断检验必须将模型置于被否决的境地，也就是说诊断检验方法必须对所有可能发生的偏离都很敏感。当然，任何诊断检验系统都不可能是非常全面的，因为在数据中常常会有一些意料之外的特征可能会被忽略。但是，如果诊断检验是经过精心设计的，且将其应用于由大量数据所拟合的模型，也并没有显示出严重偏离，那么我们在使用该模型时就会更加放心。

8.1.2 过拟合

可用于诊断检验的一种方法就是过拟合。在识别出一个确信是正确的模型后，我们事实上去拟合更复杂的模型。这样就把识别出的模型置于被否决的境地，因为复杂的模型包含了多余的参数，因而也包含唯恐出现偏离的方面.至于模型应该怎样去推广，这一问题要仔细加以考虑。特别地，与在 7.3.5 节中关于模型冗余的讨论相一致，在 ARMA 模型的两边同时追加因式是愚蠢的。如果通过分析并没有表明追加因式是必要的，但是那当然也并未证明我们的模型是正确的。一个模型只有对它施以如同圣经般的检验，才能称得上是被“证明 ”正确了的。正如 Saint Paul 在致 Thessalonians 的第一封信件中所指出的：我们所能做到的就是“去证明一切，但要牢牢守住正确的东西 ”。

过拟合的例子　作为例子，我们仍考虑某些 IBM 股票价格的数据。为进行分析，使用的数据是在本书第五部分“时间序列汇集”中列出的序列 B 。这里的序列是由 1959. 6.29—1960.6.30 这段时期[①] 的 IBM 股票价格所构成。考虑 (0,1,1) 模型

$$\nabla z_t = (1 - \theta B)a_t$$

对于 255 个有效的数据进行识别、拟合，得到 $\hat{\lambda}_0 = 1 - \hat{\theta} = 0.90$ 。

(0,1,1) 模型可以完全等价地表示为如下形式

$$\nabla z_t = \lambda_0 a_{t-1} + \nabla a_t$$

用过拟合方法考虑的推广模型为 (0,3,3) 过程：

$$\nabla^3 z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t$$

或利用 (4.3.21) 写成如下形式

$$\nabla^3 z_t = (\lambda_0 \nabla^2 + \lambda_1 \nabla + \lambda_2)a_{t-1} + \nabla^3 a_t$$

用这种特定方式推广模型的直接动机是检验由 Brown[64] 提出的建议：序列应由一个自适应的二次预报函数来作预测。在第 5 章中已经证明了：对于一个 IMA$(0,q,q)$ 过程，其最优预报函数应是一个 $q-1$ 次的自适应多项式。因此，对于前面推广的 (0,3,3) 模型来说，提前 l 期的最优预报函数是 l 的二次多项式：

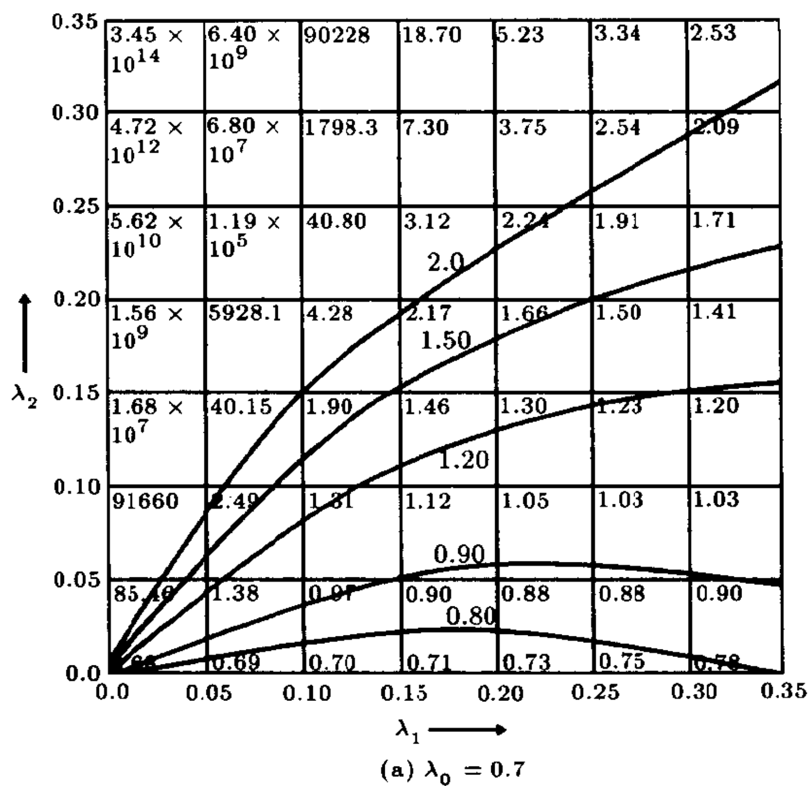
$$\hat{z}_t(i) = b_0^{(t)} + b_1^{(t)} l + b_1^{(t)} l^2$$

这里的系数 $b_0^{(t)}, b_1^{(t)}$ 和 $b_2^{(t)}$ 随着每一个新数据的获得而调整。

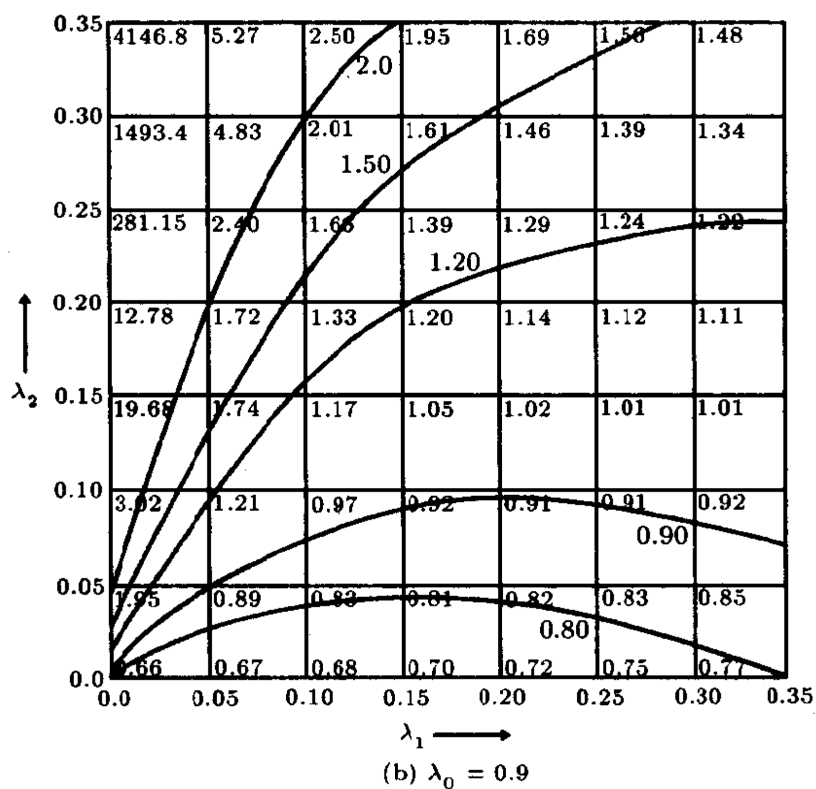
然而，我们已识别出的模型为一个 (0,1,1) 过程，得出的预报函数是

$$\hat{z}_t(l) = b_0^{(t)} \tag{8.1.1}$$

[①] 前面考虑的 IBM 股票数据也叫做序列 B，但复盖了另一段时期：1961.5.17–1962.11.2 。

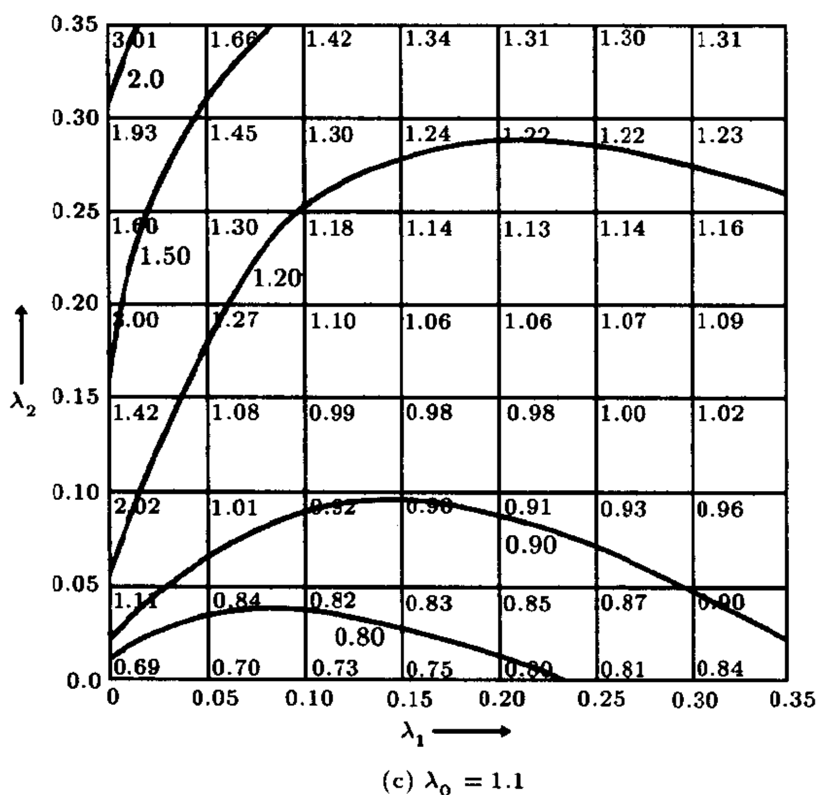


(a) $\lambda_0 = 0.7$



(b) $\lambda_0 = 0.9$

图 8.1 序列 B 关于 (0,3,3) 阶推广模型的平方和格点值和等值线



(c) $\lambda_0 = 1.1$

图 8.1 序列 B 关于 (0,3,3) 阶推广模型的平方和格点值和等值线 (续)

这是 l 的零次多项式. 因此, 隐含该模型的预报值与 l 独立, 也就是对于任何具体时刻 t 它的提前 1 期, 2 期等等的预报值都是相同的. 换句话说, 序列中只包含序列未来水平的信息, 而没有斜率或曲率的任何信息. 初看上去这似乎有些奇怪, 因为由事后的观察, 在序列中似乎存在非常确定的线性或非线性趋势. 而 λ_1 和 λ_2 非零将导致预报的趋势性, 因此检验这两个值非零到底是否发生是很有意义的. 图 8.1 给出了在 $\lambda_0 = 0.7, 0.9, 1.1$ 上 $S(\lambda_1, \lambda_2|\lambda_0)$ 的平方和格点值. 从图中可以看出最小值点接近于 $\hat{\lambda}_0 = 0.9, \hat{\lambda}_1 = 0$ 和 $\hat{\lambda}_2 = 0$, 还可以清楚地看到 λ_1 和 λ_2 的值将会导致较大的平方和值, 因此, 对识别出的 IMA(0,1,1) 模型在这两个方向上的偏离是与事实相悖的. 具体地说, 这意味着二次的预报函数将会比预报函数 (8.1.1) 给出更差而不是更好的预报值, 这事实上就是在 A5.3.3 节中已说明的情形.

8.2 应用于残差的诊断检验

过拟合方法是在某个特定的方向上对模型加以推广，并假设我们已经知道哪类偏差是要提防的。对于这种知识依赖较少的方法是建立在残差分析的基础上。然而，不能过分地强调对残差图本身的直觉观察，尽管这是在检验过程中不可缺少的第一步。

8.2.1 自相关检验

设 $w_t = \nabla^d z_t$, 假设模型

$$\phi(B)\widetilde{w}_t = \theta(B)a_t$$

已被拟合出，并得到参数的 ML 估计 $(\hat{\phi}, \hat{\theta})$ 。于是，我们称该量

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)\widetilde{w}_t \tag{8.2.1}$$

为残差。残差由 $\hat{\theta}(B)\hat{a}_t = \hat{\phi}(B)\widetilde{w}_t$ 递推算出，即为

$$\hat{a}_t = \widetilde{w}_t - \sum_{j=1}^{p}\hat{\phi}_j\widetilde{w}_{t-j} + \sum_{j=1}^{q}\hat{\theta}_j\hat{a}_{t-j} \qquad (t = 1, 2, \cdots, n)$$

有关递推时 a 和 w 的初值，或是用零初值 (条件方法), 或是用反相预报初值 (精确方法) . 于是可以证明，如果模型是适合的，有

$$\hat{a}_t = a_t + O\left(\frac{1}{\sqrt{n}}\right)$$

当序列的长度增大时， $\hat{a}_t$ 就越来越接近于白噪声 a_t 。因此，可望通过对 $\hat{a}_t$ 的研究来指示模型不当的存在与特征。特别地， $\hat{a}_t$ 自相关估计值的可辨认形态会指明对模型应做的适当修改。这一点将在 8.3 节中进一步讨论。

现在假设模型的形式是正确的，并且我们已知参数的真值。于是，利用 (2.1.13) 及 Anderson [6] 的一个结果，可知 a 的自相

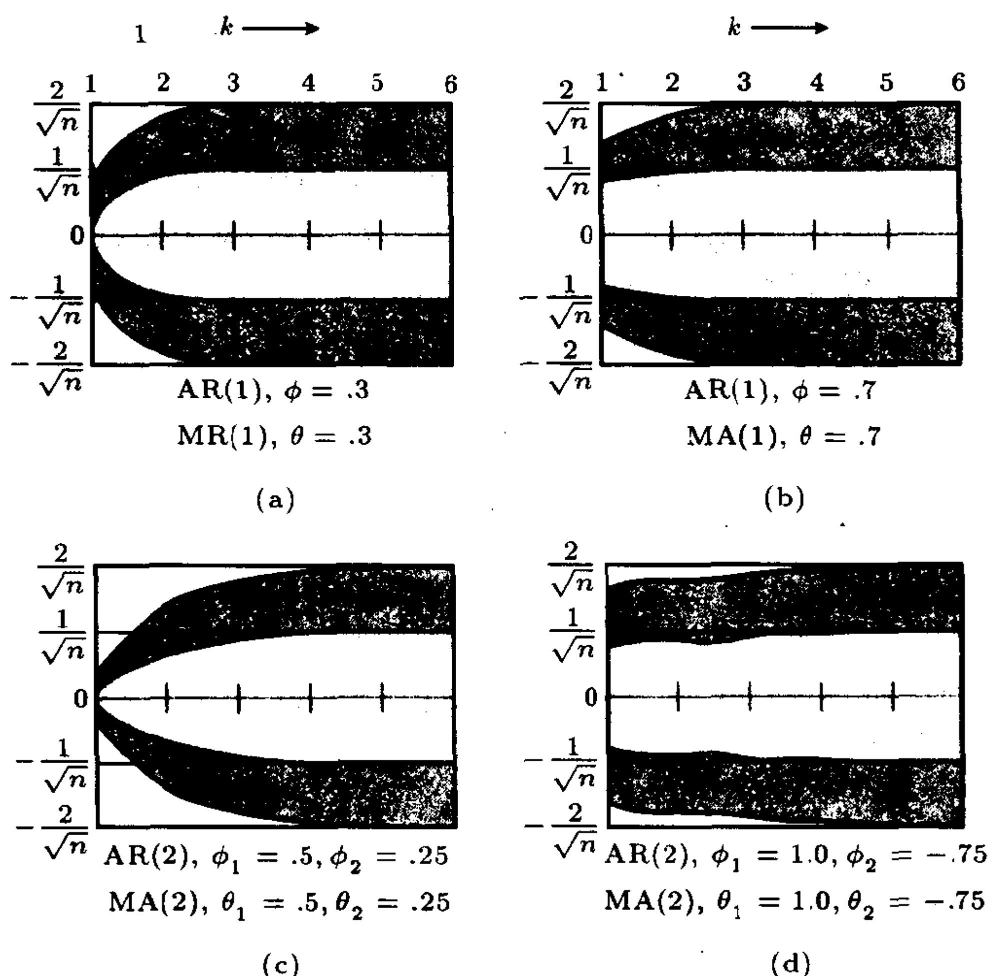


图 8.2 残差自回归 $r_k(\hat{a})$ 的标准差限

关估计 $r_k(a)$ 是互不相关的，且近似于 0 附近方差为 n^{-1} 的正态分布，从而标准差就是 $n^{-1/2}$ 利用这些事实我们可以近似地评价自相关从零点明显偏离的统计显著性。

然而，事实上我们并不知道真正的参数值，只有估计值 $(\hat{\phi},\hat{\theta})$。于是，利用 (8.2.1) 我们能够计算的不是 a 而是 $\hat{a}$ 。从 $\hat{a}_t$ 的自相关就能得出有关拟合欠缺的有用证据和模型不当的可能特征。但是，Durbin[88] 已指出：标准差 $n^{-1/2}$ 对于 $r_k(a)$ 是适合的，而在此基础上来评价自相关估计值 $r_k(\hat{a})$ 从理论值 (0) 偏离的统计显著性却可能是很危险的。Durbin 能够说明，比如对于具有参数 ϕ 的 AR(1) 过程，$r_1(\hat{a})$ 的方差是 $\phi^2 n^{-1}$，该值可能远远小于 n^{-1} 。随后，Box 和 Pierce [56] 对由任意 ARMA 过程所得 $\hat{a}$ 的自相关函数推导出了大样本的方差和协方差。他们并证明了：在任何情

况下对于低阶滞后会发生方差减小，而且在低阶滞后的 $r_k(\hat{a})$ 可能会高度相关，尽管如此，这些影响在高阶滞后会很快地消失。因此，以 $n^{-1/2}$ 作为 $r_k(\hat{a})$ 的标准差会低估低阶滞后时自相关对零明显偏离的统计显著性，但通常可以对一般和较高的滞后使用。

为了说明起见，在图 8.2 中，对于两个一阶自回归过程和两个二阶自回归过程给出了 $r_k(\hat{a})$ 大样本的一倍和二倍的标准差限。它们也对同样参数的滑动平均过程提供了相应的近似标准差 (如图 8.2 所示)。

我们可以得出结论：除非对于相当高的滞后，$n^{-1/2}$ 应被视为对 $r_k(\hat{a})$ 的标准差提供了一个上界，而不能就作为标准差本身。如果对于低阶滞后也用 $n^{-1/2}$ 作为 $r_k(\hat{a})$ 的标准差，那么我们可能严重地低估明显偏离的显著性。

8.2.2 拟合不足的一揽子检验

如果对 $r_k(\hat{a})$ 不是去单个地加以考虑，比如，将前 20 个自相关函数作为一个整体考虑，通常还需要有一项指标来指出是否存在模型不当。假设我们有任意 ARMA(p,q) 模型的前 K 个[①] 自相关 $r_k(\hat{a})(k=1,2,\cdots,K)$，那么，如果拟合模型是适合的，则可以证明 [56]

$$Q = n\sum_{k=1}^{K} r_k^2(\hat{a}) \tag{8.2.2}$$

为近似的 $\chi^2(K-p-q)$ 分布，其中 $n=N-d$ 是用于模型拟合的 w 的数目。另一方面，如果模型是不适合的，则 Q 的平均值会陡增。因此，为了考虑上面讨论的问题，设计出了一种对模型适当性假设的近似综合检验方法，也就是所谓的“一揽子”检验，该检验可以通过观测到的 Q 值并参照 χ^2 的百分点表 (如第五部分的 F 表) 而进行。

[①] 这里假设 K 取得足够大，从而记为形式 $\tilde{w}_t=\phi^{-1}(B)\theta(B)a_t=\psi(B)a_t$ 的模型中的权数 ψ_j 在 $j=K$ 之后小到可以忽略。

但是，Ljung 和 Box[138] 曾指出：在零假设之下，χ^2 分布并不能对统计量 Q 的分布提供足够的近似，Q 值比在 χ^2 分布下所预期的略为偏小. Davies, Trigges 和 Newbold[78] 提供的经验证据支持上述观点. 由此，Ljung 和 Box 给出了对上面统计量的一种修正形式 (修正的 Ljung–Box–Pierce 统计量):

$$\widetilde{Q} = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}r_k^2(\hat{a}) \tag{8.2.3}$$

该修正统计量的均值近似为 $E[\widetilde{Q}] \simeq K-p-q$, 这是 $\chi^2(K-p-q)$ 的均值。(8.2.3) 得到改进的原因在于：对于来自白噪声序列 $r_k(\hat{a})$ 的方差而言，更精确的值是 $(n-k)/n^2$, 而不是 $1/n$ 。一揽子检验统计量的修正形式被建议用于具有典型的样本容量 n, 且零假设分布非常接近于 $\chi^2(K-p-q)$ 分布的情形，当然，Davies 等人 [78] 也指出，$\widetilde{Q}$ 的方差有可能会超过相应 χ^2 分布的方差。

对序列 C 拟合的 (0,2,2) 模型　为了举例说明一揽子检验准则 (8.2.3), 在表 8.1 中列出了由 IMA(0,2,2) 过程 $\nabla^2 z_t = (1-0.13B-0.12B^2)a_t$ 得出自相关 $r_k(\hat{a})$ 的前 25 个值，(0,2,2) 是在第 7 章中对序列 C 拟合的模型之一. 因为有 $n=224$ 个 w, 所以单个自相关的标准差近似上界为 $1/\sqrt{224} \simeq 0.07$ 。与该标准差的界限相比，$r_3(\hat{a}) = -0.125, r_9(\hat{a}) = -0.130, r_{11}(\hat{a}) = -0.219, r_{17}(\hat{a}) = 0.153, r_{22}(\hat{a}) = 0.132$, 以及 $r_{25}(\hat{a}) = -0.127$, 这些值都是相当大的。当然，即便在随机序列中偶尔也会出现大的偏差，但是把这

表 8.1　对序列 C 拟合模型 $\nabla^2 z_t = =(1-0.13B-0.12B^2)a_t$ 所得残差的自相关 $r_k(\hat{a})$

k	$r_k(\hat{a})$	k	$r_k(\hat{a})$	k	$r_k(\hat{a})$	k	$r_k(\hat{a})$	k	$r_k(\hat{a})$
1	0.020	6	−0.033	11	−0.129	16	−0.050	21	0.007
2	0.032	6	0.022	12	0.063	17	0.153	22	0.132
3	−0.125	8	−0.056	13	−0.084	18	−0.092	23	0.012
4	−0.078	9	−0.013	14	0.022	19	−0.005	24	−0.012
5	−0.011	10	0.093	15	−0.006	20	−0.015	25	−0.127

表 8.2　对序列 C 拟合模型 $(1-0.82B)\nabla z_t$ 所得残差的自相关 $r_k(\hat{a})$

k	$r_k(\hat{a})$	k	$r_k(\hat{a})$	k	$r_k(\hat{a})$	k	$r_k(\hat{a})$	k	$r_k(\hat{a})$
1	−0.007	6	0.019	11	−0.098	16	−0.039	21	0.001
2	−0.002	6	0.073	12	0.074	17	0.165	22	0.129
3	−0.061	8	−0.030	13	−0.054	18	−0.083	23	0.014
4	−0.014	9	−0.097	14	0.034	19	−0.004	24	−0.017
5	0.047	10	0.133	15	0.002	20	−0.009	25	−0.129

些结果作为一个整体考虑，就必定会有模型拟合不当的嫌疑。

为了作更正式的评价，我们对

$$\widetilde{Q}=(224)(226)\left[\frac{(0.020)^2}{223}+\frac{(0.032)^2}{222}+\cdots+\frac{(-0.127)^2}{199}\right]=36.2$$

参照自由度为 23 的 χ^2 表。自由度为 23 的 χ^2 分布的 10% 和 5% 点分别是 32.0 和 35.2. 因此，该模型的恰当性存在一些疑问。

对序列 C 拟合的 (1,1,0) 模型　我们在第 7 章中确定的模型 $(1-0.82B)\nabla z_t=a_t$ 对序列 C 给出了一个更好的描述，表 8.2 中给出了该模型残差的前 25 个自相关。对于该模型有 $\widetilde{Q}=(225)(227)\sum_{k=1}^{25} r_k^2(\hat{a})/(225-k)=31.3$ 。参照自由度为 24 的 χ^2-表，说明没有理由对此模型质疑。

基于 $K=25$ 个残差自相关，表 8.3 对于表 7.11 中拟合序列 A 到 F 的模型概括了准则 $\widetilde{Q}$ 的值。但是，特别对于长度较短的序列，如序列 E 和 F, 建议在 (8.2.3) 中实际采用小一些的 K 值，因为随着样本长度 n 的增加，涉及统计量 $\widetilde{Q}$ 分布的渐近理论也依赖于 K 的增加 (但只是缓慢的，满足 $K/n\to 0$) (见 8.2.4 节末尾的有关讨论)。

考察表 8.3 可见，只出现两个 $\widetilde{Q}$ 较大的可疑值，一个是在对序列 C 拟合 (0,2,2) 模型后得到的值 36.2, 对此我们已作过讨论。另一个是对序列 B 拟合 (0,1,1) 模型后得到的值 $\widetilde{Q}=38.8$ 。由于

表 8.3 对序列 A 到 F 拟合各种模型所得残差做残差一揽子检验的总结

序列	$n=N-d$	拟合模型	$\tilde{Q}$	自由度
A	197	$z_t-0.92z_{t-1}=1.45+a_t-0.58a_{t-1}$	28.4	23
	196	$\nabla z_t=a_t-0.70a_{t-1}$	31.9	24
B	368	$\nabla z_t=a_t-0.09a_{t-1}$	38.8	24
C	225	$\nabla z_t-0.82\nabla z_{t-1}=a_t$	31.3	24
	224	$\nabla^2 z_t=a_t-0.13a_{t-1}-0.12a_{t-2}$	36.2	23
D	310	$z_t-0.87z_{t-1}=1.17+a_t$	11.5	24
	309	$\nabla z_t=a_t-0.06a_{t-1}$	18.8	24
E	100	$z_t-1.42z_{t-1}+0.37z_{t-2}=14.35+a_t$	26.8	23
	100	$z_t-1.57z_{t-1}+1.02z_{t-2}-0.21z_{t-3}=11.31+a_t$	20.0	22
F	70	$z_t+0.34z_{t-1}-0.19z_{t-2}=58.87+a_t$	14.7	23

自由度为 24 的 χ^2 分布 5% 和 10% 点分别为 36.4 和 39.3, 这就意味着某些模型是不恰当的。我们在下一节中将讨论这种不当的可能特征。

8.2.3 参数值改变引起的模型不当

可以想象, 若模型形式保持不变, 但在很长的时间范围内模型的参数改变了, 由此而发生的是模型不当的一种重要形式。对于 IBM 股票拟合的 (0,1,1) 模型, 用这种方式来解释模型可能的不当是有根据的。

表 8.4 给出了分别对序列 B 的前半部分、后半部分以及整个序列拟合 (0,1,1) 过程所得的结果. 用 $\hat{\lambda}^{(1)}$ 和 $\hat{\lambda}^{(2)}$ 表示从两个部分所得的估计值, 我们发现 $\hat{\lambda}^{(1)}-\hat{\lambda}^{(2)}$ 的标准差是 $\sqrt{(0.070)^2+(0.074)^2}=0.102$ 。因为差值为 $\hat{\lambda}^{(1)}-\hat{\lambda}^{(2)}=0.26$, 它是标准差的 2.6 倍, 所以看来很可能 λ 真的发生了变化。对 $\tilde{Q}$ 的考察表明, 对于不同时间段的参数作适当修改后的 (0,1,1) 模型可以更确切地解释这个序列。残差方差 $\hat{\sigma}_a^2$ 的估计结果也显著表明, 序列两部分之间的波动程度确实发生了变化。根据对数据的更进一步考察, 序列特征的变化大约发生在时刻 $t=236$ 附近。

表 8.4 对序列 B 的前半部和后半部拟合 (0,1,1) 模型对比

	n	$\hat{\theta}$	$\hat{\lambda}=1-\hat{\theta}$	$\hat{\sigma}(\lambda)=\left[\frac{\hat{\lambda}(2-\hat{\lambda})}{n}\right]^{1/2}$	残差方差 $\hat{\sigma}_a^2$	$\hat{Q}$	自由度
前半部	184	−0.29	1.29	±0.070	26.3	24.6	24
后半部	183	−0.03	1.03	±0.074	77.3	37.1	24
整体	368	−0.09	1.09	±0.052	52.2	38.8	24

8.2.4 用于模型检验的得分检验

作为模型检验中直接使用过拟合方法的替代,还可以使用 Lagrange 乘子检验或得分检验方法,这种方法与一揽子检验方法联系紧密。一般的比值检验方法由 Silvey[179] 给出,该方法在 ARIMA 模型中的应用 Godfrey[96],Poskitt 和 Tremayne[159] 都曾讨论过。得分检验在计算上的优势在于它仅要求被检验的零模型中参数的极大似然估计,但得到的检验却渐近等价于直接由过拟合模型得到的似然比检验。而且该检验的统计量很容易计算,它的形式为样本数目 n 乘以由特定“辅助”回归决定的一个系数。

因此,假设已对观测 $\tilde{w}_t$ 用 ML 估计了一个 ARMA(p,q) 模型,通过检验零模型与对立的备择模型 ARMA$(p+r,q)$ 或 ARMA$(p,q+r)$,我们希望评价模型的恰当性。即对于备择模型 ARMA$(p+r,q)$,检验 $H_0:\phi_{p+1}=\cdots=\phi_{p+r}=0$,而对备择模型 ARMA$(p,q+r)$,检验 $H_0:\theta_{q+1}=\cdots=\theta_{q+r}=0$ 。得分检验方法是建立在对数似然函数一阶偏导数 (或对数似然函数的得分) 的基础上的,似然函数关于备择模型进行计算,但它是在零模型下所得 ML 估计处取值。对数似然函数通常给出为 $l=-(n/2)\ln(\sigma_a^2)-(1/2\sigma_a^2)\sum_{t=1}^{n}a_t^2$ 。于是, l 关于参数 $(\boldsymbol{\phi},\boldsymbol{\theta})$ 的偏导数为

$$\frac{\partial l}{\partial \phi_j}=-\frac{1}{\sigma_a^2}\sum_{t=1}^{n}\frac{\partial a_t}{\partial \phi_j}a_t$$

$$\frac{\partial l}{\partial \theta_j}=-\frac{1}{\sigma_a^2}\sum_{t=1}^{n}\frac{\partial a_t}{\partial \theta_j}a_t$$

与 (7.2.9) 和 (7.2.10) 中一样，我们有

$$-\frac{\partial a_t}{\partial \phi_j} = u_{t-j}\ , \qquad -\frac{\partial a_t}{\partial \theta_j} = v_{t-j}$$

这里， $u_t = \theta^{-1}(B)\widetilde{w}_t = \phi^{-1}(B)a_t$, 及 $v_t = -\theta^{-1}(B)a_t$ 。对零模型进行 ML 拟合得到的残差 $\hat{a}_t$ 给出如下：

$$\hat{a}_t = \widetilde{w}_t - \sum_{j=1}^{p}\hat{\phi}_j\widetilde{w}_{t-j} + \sum_{j=1}^{q}\hat{\theta}_j\hat{a}_{t-j} \quad (t = 1, 2, \cdots, n)$$

在零模型的 ML 估计下取值的 u_t 与 v_t 可以递推地计算，例如初值可取为零，递推格式为

$$\begin{aligned} u_t &= \widetilde{w}_t + \hat{\theta}_1 u_{t-1} + \cdots + \hat{\theta}_q u_{t-q} \\ v_t &= -\widetilde{a}_t + \hat{\theta}_1 v_{t-1} + \cdots + \hat{\theta}_q v_{t-q} \end{aligned}$$

关于所有模型参数一阶偏导数的得分向量可表示为

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma_a^2}\boldsymbol{X}'\boldsymbol{a} \tag{8.2.4}$$

这里，$\boldsymbol{a} = (a_1, \cdots, a_n)'$,$\boldsymbol{X}$ 是 $n\times(p+q+r)$ 阶矩阵，在 ARMA$(p+r, q)$ 备择的情形，该矩阵的第 t 行是 $(u_{t-1},\cdots,u_{t-p-r},v_{t-1},\cdots,v_{t-q})$，而在 ARMA$(p,q+r)$ 备择的情形，第 t 行是 $(u_{t-1},\cdots,u_{t-p},v_{t-1},\cdots,v_{t-q-r})$ 。于是，与 (7.2.17) 类似，由于 $\boldsymbol{\beta}$ 的大样本信息阵可以由 $\hat{\sigma}_a^{-2}\boldsymbol{X}'\boldsymbol{X}$ 一致地估计，这里 $\hat{\sigma}_a^2 = n^{-1}\sum_{t=1}^{n}\hat{a}_a^2 = n^{-1}\hat{\boldsymbol{a}}'\hat{\boldsymbol{a}}$, 因此，检验附加的 r 个参数等于 0 的得分检验统计量为

$$\Lambda = \frac{\hat{\boldsymbol{a}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{a}}}{\hat{\sigma}_a^2} \tag{8.2.5}$$

Godfrey[96] 指出，(8.2.5) 中统计量的计算可以解释为等于由辅助回归方程中确定出的系数乘以 n 倍，也就是，若备择模型为 ARMA$(p+r,q)$, 我们考虑辅助回归方程

$$\hat{a}_t = \alpha_1 u_{t-1} + \cdots + \alpha_{p+r}u_{t-p-r} + \beta_1 v_{t-1} + \cdots + \beta_q v_{t-q} + \epsilon_t$$

而当备择模型为 ARMA$(p,q+r)$ 时，我们考虑回归方程

$$\hat{a}_t = \alpha_1 u_{t-1} + \cdots + \alpha_p u_{t-p} + \beta_1 v_{t-1} + \cdots + \beta_{q+r} v_{t-q-r} + \epsilon_t$$

用 $\hat{\epsilon}_t$ 表示该回归方程普通最小二乘估计的残差，那么，由 (8.2.5) 看出，Λ 实质上可以表示为

$$\Lambda = \frac{n(\sum_{t=1}^{n}\hat{a}_t^2 - \sum_{t=1}^{n}\hat{\epsilon}_t^2)}{\sum_{t=1}^{n}\hat{a}_t^2} = n\big(1 - \frac{\sum_{t=1}^{n}\hat{\epsilon}_t^2}{\sum_{t=1}^{n}\hat{a}_t^2}\big)$$

它是 $\hat{a}_t$ 对 v_{t-j} 和 u_{t-j} 作回归所确定系数的 n 倍。在拟合 ARMA(p, q) 模型是正确的零假设之下，统计量 Λ 具有自由度为 r 的渐近 χ^2 分布，对于大的 Λ 值，零模型被拒绝并认其为不当. 正如 Godfrey[96] 和其它人所指出的，若通过得分检验方法对零模型拒绝，但这并不能作为采纳具体备择模型的依据，它只能作为拟合模型不恰当的证据。与此相应地，可以感到应该有一种推广形式的得分检验方法，它具有相当的效力去检验拟合模型中的不恰当性，甚至当正确的模型不是备择假设下的特定形式，这种检验方法仍然有效。特别地，例如 Poskitt 和 Tremayne[159] 证明了，针对备择模型 ARMA$(p+r,q)$ 的得分检验渐近等价于针对备择模型 ARMA$(p,q+r)$ 的检验。因此，得分检验方法对于备择假设下特定的具体模型是不敏感的，但是，该检验的实现当然会依赖于特定附加参数个数 r 的选择。

我们还注意到得分统计量 Λ 的另一种形式。利用 ML 估计方法，可知一阶偏导数 $\partial l/\partial\phi_j, j=1,\cdots,p$ 及 $\partial l/\partial\theta_j, j=1,\cdots,q$ 在 ML 估计值处取值皆为零，因此，得分向量 $\partial l/\partial\boldsymbol{\beta}$ 在零模型的 ML 估计处取值时，只包含 r 个非零元素，它们是关于备择模型 r 个附加参数的偏导数。因此 (8.2.5) 中的得分统计量也可被看作是这 r 个非零值的二次型，当在零模型下所得 ML 估计处取值时，这个二次型的矩阵是这 r 个得分值协方差阵之逆的相合估计。由

于这 r 个得分值在零模型下是均值为零的渐近正态分布，因此容易看出，在零假设下渐近 $\chi^2(r)$ 分布也是正确的。Newbold[150] 还指出，针对 r 个附加参数备择假设的得分统计检验方法，与建立在前 r 个拟合模型残差自相关 $r_k(\hat{a})$ 之上的适当统计量是密切相关的。在前 r 个残差自相关中，检验统计量基本是二次形式，但是，比 (8.2.2) 中一揽子检验统计量的形式要复杂一些。作为简明的示例，假设拟合模型和零模型是纯 AR(p) 模型，而在得分检验方法中的备择模型是 ARMA(p,r) 模型。于是由前述可得变量 v_{t-j} 等于 $-\hat{a}_{t-j}$，这是因为在零模型假设下 $\theta(B)\equiv 1$。因此，(8.2.4) 中得分向量的非零元素等于 $-n$ 乘以拟合模型的前 r 个残差自相关 $r_1(\hat{a}),\cdots,r_r(\hat{a})$，从而直接可以看出得分检验统计量就是这前 r 个残差自相关的二次型。鉴于与其它因素之间的联系，Ljung[137] 建议当用于模型检验时，在 (8.2.3) 中修正的一揽子检验统计量应采用相对较小的 K 值 (例如 $K=5$)。

8.2.5 累积周期图检验

在某些场合，特别是在第 9 章所讨论的季节时间序列拟合中，恐怕很可能会未充分考虑到序列的周期特性，因此，我们应注意残差中的周期性。自相关函数对于这类从随机状态的偏离并不能给出灵敏的指示，因为周期效应本身常常融会在自相关之中。而另一方面，周期图就是为检验在白噪声背景下周期波形的模式而设计的。

正如 2.2 节中所定义的，一个时间序列 $a_t, t=1,2,\cdots,n$ 的周期图是

$$I(f_i)=\frac{2}{n}\left[\left(\sum_{t=1}^{n}a_t\cos(2\pi f_i t)\right)^2+\left(\sum_{t=1}^{n}a_t\sin(2\pi f_i t)\right)^2\right] \tag{8.2.6}$$

其中 $f_i=i/n$ 为频率。因此周期图是把 a_t 和不同频率的正弦和余弦波相联系的一种工具。在残差中若含给定的频率，则在该频率上相应的正弦或余弦波会使波形增强，从而产生大的 $I(f_i)$ 值。

累积周期图　Bartlett[26] 已经证明了 (也可见 [122]), 累积周期图为非随机的周期检测提供了有效的工具。白噪声的功率谱 $p(f)$ 在 0 到 0.5 周的频率区域上都具有常值 $2\sigma_a^2$ 。因此，白噪声的累积功率谱

$$P(f) = \int_0^f p(g)dg \tag{8.2.7}$$

对 f 作图就是从 (0,0) 到 $(0.5,\sigma_a^2)$ 的直线，即 $P(f)/\sigma_a^2$ 是一条从 (0,0) 到 (0.5,1) 的直线。如 2.2.3 节所述，$I(f)$ 给出了频率 f 处功率谱的估计。事实上，对于白噪声有 $E[I(f)] = 2\sigma_a^2$, 因此估计是无偏的。故 $(1/n)\sum_{i=1}^{j}I(f_i)$ 给出了积分功率谱的无偏估计，且

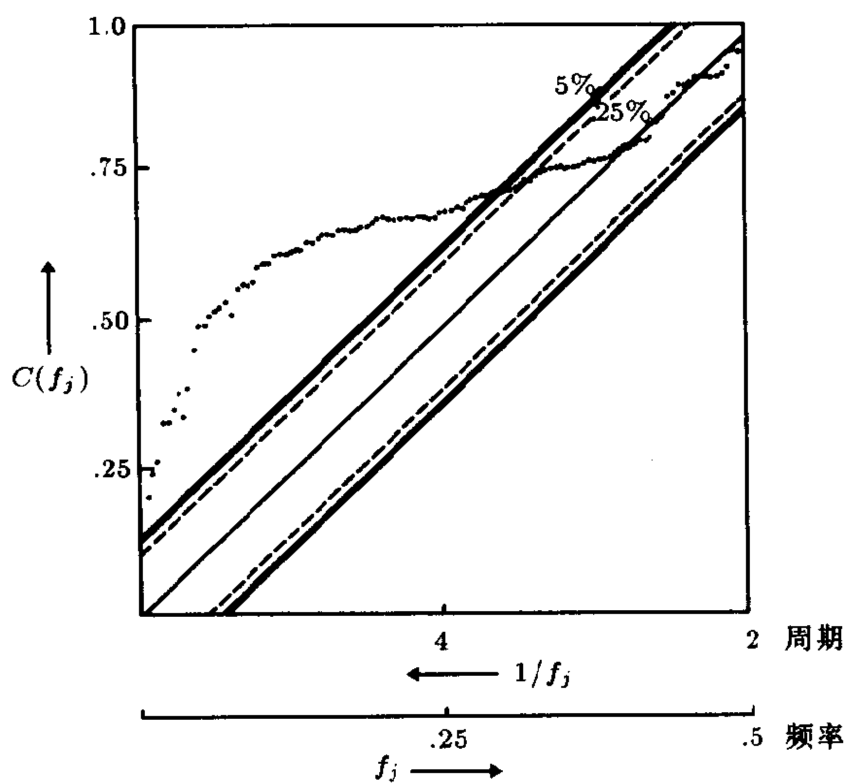
$$C(f_j) = (ns^2)^{-1}\Big(\sum_{i=1}^{j} I(f_j)\Big) \tag{8.2.8}$$

是 $P(f_j)/\sigma_a^2$ 的估计，其中 s^2 是 σ_a^2 的估计。我们称 $C(f_j)$ 为归一化的累积周期图。

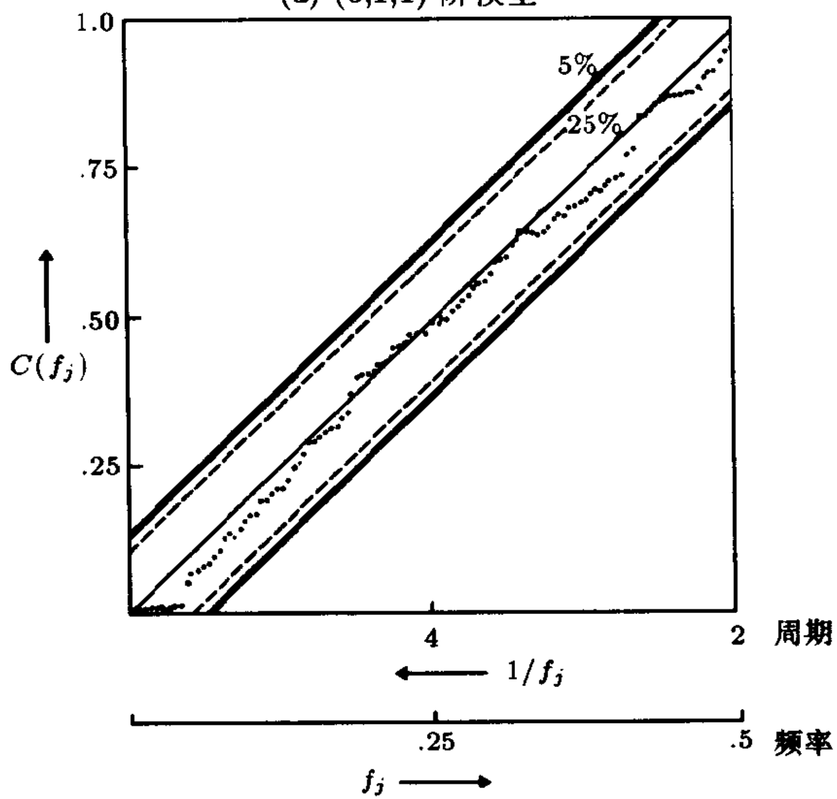
现在，如果模型是恰当的且参数精确地已知，那么，a 就可以从数据算出，并得到一个白噪声序列。对于白噪声序列来说，$C(f_j)$ 对于 f_j 的图将会散布在连接点 (0,0) 和 (0.5,1) 的直线附近。另一方面，模型不恰当将会产生非随机的 a, 累积周期图就会表现出对上述直线的系统偏离。特别地，a 的周期性会导致 $I(f_j)$ 产生一系列相邻的大值。这些大的纵坐标值在 $C(f_j)$ 中相互加强，从而相对于预期的直线形成一个隆起。

事实上，我们并不知道参数的精确值，仅知道其估计值。我们没有 a, 只有估计的残差 $\hat{a}$ 。然而对于大样本来说，$\hat{a}$ 和 a 的周期图具有相似的特征，因此，仔细考察 $\hat{a}$ 的周期图能够提供有用的补充诊断检验，这种方法特别可以指出考虑不当的周期性。

例：序列 C　我们已经知道序列 C 用下面的 (1,1,0) 模型可



(a) (0,1,1) 阶模型



(b) (0,2,2) 阶模型

图 8.3　序列 C: 最佳拟合过程残差的累积周期图

以很好地拟合：

$$(1-0.82B)\nabla z_t = a_t$$

而用 IMA(0,2,2) 模型

$$\nabla^2 z_t = (1-0.13B-0.12B^2)a_t$$

就稍差一些，但它与前一模型几乎是等价的。我们对序列 C 拟合不恰当的 IMA(0,1,1) 模型

$$\nabla z_t = (1-\theta B)z_t$$

然后分析残差 a_t 看到底会发生什么，由此来说明累积周期图检验。

当模型被如此限定时，我们求出 θ 的最小二乘估计是 –0.65. 该模型残差的归一化累积周期图在图 8.3(a) 中给出。通过对图形的考察显然可知，累积周期图呈现出对线性的明显偏离。这些偏离在低频处非常明显，在本例中由于差分不足出现这种情况是可以料想的。图 8.3(b) 对拟合最好的 IMA(0,2,2) 过程给出了相应的图形。这时，累积周期图的点更紧密地聚集在期望直线附近，但是，正如我们在图 8.3(b) 中所见，还有其它迹象显示出该模型的不当之处。

在图上既标明周期也标明频率，这是很方便的，这样是为了当残差包含周期而出现隆起时便于识别。例如在月度销售数据中，隆起靠近周期 12,24,36 等等，这表明对季节影响考虑不当。

累积周期图和积分谱之间的概率关系与经验累积频率函数和累积分布函数之间的关系完全一样。正因如此，我们可以利用 Kolmogorov-Smirnov 检验 [99] 判断周期图对于预期 $\hat{a}_t$ 为白噪声情形的偏离。利用这一检验，我们可以在理论直线附近设置概率限，若序列 $\hat{a}_t$ 为白噪声时，周期图相对于理论直线的偏离，仅以指定概率越过此界限。然而， $\hat{a}_t$ 是拟合值，而非真值 a_t，故可知即使模型是正确的， $\hat{a}_t$ 也不是精确的白噪声过程。因此，作为

模型不恰当性的检验， Kolmogorov-Smirnov 概率限的应用只能指出近似的概率值。然而，给出累积周期图上的概率限是有意义的，因为它提供了粗略的指向 — 什么样的偏离应视为可疑，什么样的偏离要进一步注意。

对于真正随机的序列，上述概率限将会以时间的比例 ϵ 被越过。它们画在理论直线上、下方距离为 $\pm K_\epsilon/\sqrt{q}$ 处，这里，若 n 为偶数， $q=(n-2)/2$; 若 n 为奇数， $q=(n-1)/2$ 。表 8.5 给出了 K_ϵ 的近似值。

表 8.5 对于累积周期图计算近似概率限的系数

ϵ	0.01	0.05	0.10	0.25
K_ϵ	1.63	1.36	1.22	1.02

对于序列 C 来说， $q=(224-2)/2=111$, 在图 8.3 中嵌上的 5% 概率限从理论直线的偏离值 $\pm 1.36\sqrt{111}=\pm 0.13$ 。与此类似， 25% 概率限的偏离值为 $\pm 1.02\sqrt{111}=\pm 0.10$ 。

结论 前文所述检验方法都有各自的优点和缺点。建立在考察自相关函数估计值和累积周期图基础上的检验方法，虽然能够指出序列不容置疑的特性，但可能并不十分灵敏。过拟合方法对于特定偏差的检验更灵敏一些，但它无法提防一些意外的麻烦。

8.3 利用残差修正模型

8.3.1 使用不正确模型时残差的相关特征

当某个拟合模型的残差自相关函数表明模型不恰当时，就必须考虑应该如何对模型进行修正。在 8.3.2 节中我们说明了如何利用残差自相关给出修正方案。作为介绍，我们考虑拟合不正确的模型对残差自相关的影响。

假设正确的模型是

$$\phi(B)\widetilde{w}_t=\theta(B)a_t$$

而所采用的不正确模型是

$$\phi_0(B)\widetilde{w}_t = \theta_0(B)b_t$$

于是不正确模型的残差 b_t 是相关的，并且因为

$$b_t = \theta_0^{-1}(B)\theta(B)\phi_0(B)\phi^{-1}(B)a_t \tag{8.3.1}$$

所以 b 的自协方差生成函数为

$$\sigma_a^2[\theta_0^{-1}(B)\theta_0^{-1}(F)\theta(B)\theta(F)\phi_0(B)\phi_0(F)\phi^{-1}(B)\phi^{-1}(F)] \tag{8.3.2}$$

例如，假设在 IMA(0,1,1) 模型中，我们用某个其它的 θ_0 值来代替正确的 θ 值。那么，残差将服从下面的 (1,0,1) 阶混合过程

$$(1-\theta_0 B)b_t = (1-\theta B)a_t$$

利用 (3.4.8), 我们有

$$\rho_1 = \frac{(1-\theta\theta_0)(\theta_0-\theta)}{1+\theta^2-2\theta\theta_0}$$
$$\rho_j = \rho_1\theta_0^{j-1} \qquad (j=2,3,\cdots)$$

例如，假设在 IMA(0,1,1) 过程中，

$$\nabla z_t = (1-\theta B)a_t$$

当正确值为 $\theta=0$ 时，我们取 $\theta_0 = 0.8$ 。于是有

$$\theta_0 = 0.8 \quad , \qquad \theta = 0.0$$
$$\rho_1 = 0.8 \quad , \qquad \rho_j = 0.8^j$$

因此， b 是高度相关的，且满足自回归过程

$$(1-0.8B)b_t = a_t$$

8.3.2 利用残差修正模型

假设，来自模型

$$\phi_0(B)\nabla^{d_0}z_t = \theta_0(B)b_t \tag{8.3.3}$$

的残差 b_t 显示出非随机性。利用 b_t 的自相关函数，则可用第 6 章的方法来对序列 b_t 识别模型

$$\overline{\phi}(B)\nabla^{\overline{d}}b_t = \overline{\theta}(B)a_t \tag{8.3.4}$$

用 (8.3.3) 和 (8.3.4) 消去 b_t, 我们得出一个新的模型

$$\phi_0(B)\overline{\phi}(B)\nabla^{d_0}\nabla^{\overline{d}}z_t = \theta_0(B)\overline{\theta}(B)a_t \tag{8.3.5}$$

现在即可对该模型进行拟合及诊断检验。

例如，一个序列被错误地识别为 IMA(0,1,1) 过程，经拟合给出模型

$$\nabla z_t = (1+0.6B)b_t \tag{8.3.6}$$

再假设对残差序列识别出模型

$$\nabla b_t = (1-0.8B)a_t \tag{8.3.7}$$

那么，由 (8.3.6) 和 (8.3.7) 消去 b_t, 我们将得到

$$\nabla^2 z_t = (1-0.2B-0.48B^2)a_t$$

这表明现在我们应该接受一个 IMA(0,2,2) 过程。

9

季节模型

在第 3 章到第 8 章中，我们考虑了一类线性随机模型的性质，它们对于描述平稳和非平稳时间序列是有意义的，并且我们还看到了怎样用这些模型来预报。随之，我们考虑了将这些模型与实际数据相联系时产生的识别、拟合及诊断检验等实际问题。在这一章中，我们将运用这些方法来分析和预报季节序列，同时也提供了一个机会，使我们得以说明如何将前几章的方法运用于当前。

9.1 季节时间序列的简约模型

图 9.1 给出了 1952,1953 和 1954 年的国际航线旅客总数。这是 Brown[64] 引用的一个长序列 (12 年的数据) 中的一部分，并作为序列 G 列在本书第五部分的“时间序列汇集”中。该序列呈现明显的季节模式，因为旅行在夏季的月份达到最高峰，而第二高峰出现在春季. 许多其它序列，特别是销售数据，也呈现出类似的季节特征。

通常，在一个序列中，若经过 s 个基本时间间隔后呈现出相似性，我们就说序列表现出以 s 为周期的周期特性。在前面的例子中，基本时间间隔为 1 个月，周期为 12 个月。当然，也会出现 s 取其它值的例子。例如，对于在一年中显示出季节影响的季度

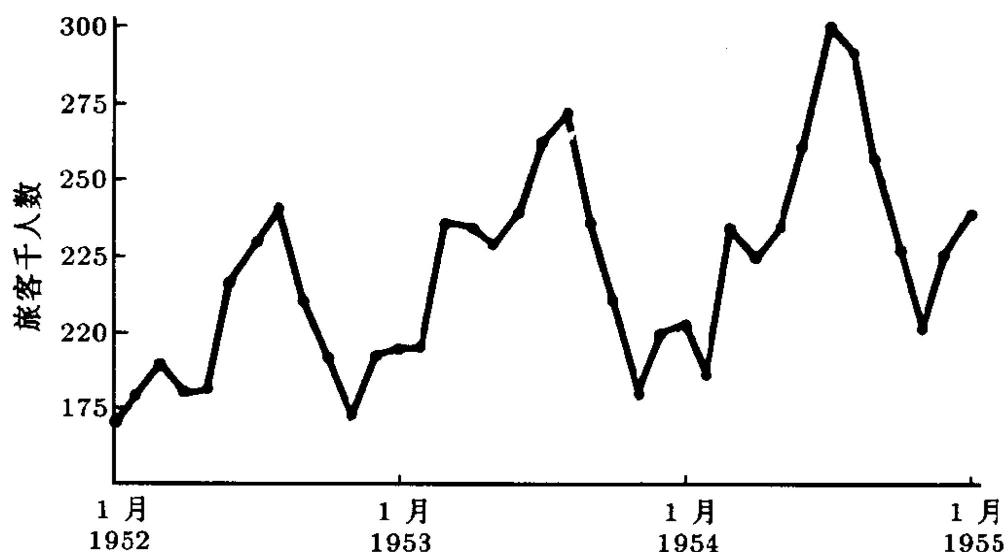


图 9.1 国际航线旅客总数 (是序列 G 的一部分，单位：千人)

数据，有 $s=4$ 。有时也可能发生不止一个周期的情况。下面就是这种情形：由于帐单往往是按月支付的，所以我们可以料想，一个银行每周所做的业务会在月度内显示出 $s=4$ 的周期，但月度业务会显示出 $s=12$ 的周期。

9.1.1 拟合与预报

以往对时间序列分析中的一个缺陷就是：对一个时间序列进行拟合与对其进行预报二者之间的混淆。例如，假设一个时间序列在某个特定时期呈现增长趋势，同时也服从某种季节模式。一种通常的分析方法就是主观地把这个序列分为三个分量："趋势分量"，"季节分量"和"随机分量"。趋势可以用一个多项式来拟合，季节分量用 Fourier 拟合。然后，进行预报是用这些拟合函数推算出的。

这种方法可能会给出极其错误的结果。例如，我们已经知道 IBM 股票价格 (序列 B) 的特征可由随机游动模型 $\nabla z_t = a_t$ 来很好地近似，即为

$$z_t = z_0 + \sum_{j=0}^{t-1} a_{t-j} \qquad (9.1.1)$$

这意味着 $\hat{z}_t(l)=z_t$ 。换言之，股票未来价格最好的预报值非常接近于今天的价格。从序列的短期数据来看，似乎它们确实可以用二次曲线来拟合。但这只是反映了随机偏离有时确实会表现出的性质。然而，使用二次的预报函数是没有根据的，这将会得出很坏的预报值。当然，那些可以作出物理解释的真正的系统影响，应该通过在模型中引入一个恰当的确定性分量而加以考虑。例如，假若已知热量被稳定地加到一个系统上，那么，就可以在随机分量之外增加一个恰当的时间确定性函数，以此来明确地解释因加热而导致的温度升高。

9.1.2 包含自适应正弦、余弦项的季节模型

通常的线性模型为

$$\widetilde{z}_t=\sum_{j=1}^{\infty}\pi_j\widetilde{z}_{t-j}+a_t=\sum_{j=1}^{\infty}\psi_j a_{t-j}+a_t \tag{9.1.2}$$

系数 π_j 和 ψ_j 取恰当的值就完全足以描述许多季节时间序列。问题在于要为这种模型选择一个简约参数的适当系统。正如我们前面提到的，这并不是一个数学问题，而是如何更好地表现世务的问题。为了做到这一点，我们只能在实际的时间序列上试验一些方法，并发展那些看来是富有成效的概念。

我们已经看到，对于非季节性的时间序列，可以得到一个有用且简约的表达式，其形式为

$$\varphi(B)\widetilde{z}_t=\theta(B)a_t \tag{9.1.3}$$

此外，广义自回归算子 $\varphi(B)$ 决定了最终预报函数，它是如下差分方程的解：

$$\varphi(B)\hat{z}_t(l)=0$$

这里， B 理解为作用在 l 上的运算。在描述季节特性时，我们想为预报函数找出一种季节模式。我们首先想到的就是 $\varphi(B)$ 可以

生成一个这样的预报函数，它由正弦和余弦混合组成，并且还可能混有多项式，因而容许序列水平的改变以及周期模式的变化。这样的预报函数能够完全自然地出现在一般模型 (9.1.3) 的结构中。例如，对于月度数据，某个预报函数是以 12 个月为周期的正弦波，相位和振幅度是适当的，预报函数将满足差分方程

$$(1-\sqrt{3}B+B^2)\hat{z}_t(l)=0$$

这里，将 B 理解为作用在 l 上的运算。然而，周期特性并不一定能用正弦和余弦的混合作出简约的表示。例如，为了描述受圣诞节，复活节以及其它季节性购买的销售额，将需要许多项正弦 - 余弦分量。考虑一个极端情形：英国的烟花销售量很大程度局限在 11 月 5 日之前的几个星期，因为这时要庆祝 Guy Fawkes 爆炸国会楼企图的失败。打算直接用正弦和余弦来描述烟花销售量的“单峰”可能是无益的，对这一问题显然需要更仔细地考虑。

从先前的分析中可知，我们没有必要去估计 $\varphi(B)$ 的所有分量。当需要 d 次差分以达到平稳时，我们记为 $\varphi(B)=\phi(B)(1-B)^d$，这相当于设定方程 $\varphi(B)=0$ 的 d 个根等于 1. 若这样的表达式被证明是合适时，我们就可以接下去对 $w_t=\nabla^d z_t$ 做较简单的分析。这样，我们使用了 $\nabla=1-B$ 作为简化算子。在其它问题中，可能不同种类的简化算子是适用的。例如，供热的燃油耗量极大地依赖于周围环境的温度，由于地球围绕太阳旋转，因此可知环境温度近似地服从周期为 12 个月的正弦波。在分析燃料油的销售量时，引入 $1-\sqrt{3}B+B^2$ 作为简化算子可能是合理的，该算子构成了广义自回归算子 $\varphi(B)$ 有供献的一个分量。如果这样的表达式被证明是有效的，那么，我们接下来就可以对 $w_t=(1-\sqrt{3}B+B^2)z_t$ 做简单的分析。我们注意到，这个算子是齐次非平稳的变形，且在单为圆上有零点 $e^{\pm(i2\pi/12)}$ 。

9.1.3 一般的乘积季节模型

简化算子 $1-B^s$　关于周期为 s 的季节时间序列有一个基本

事实，那就是时间相隔为 s 的观测值是相似的。因此，可以指望算子 $B^s z_t = z_{t-s}$ 将会在季节序列的分析中起重要作用，而且，因为在序列 $z_t, z_{t-s}, z_{t-2s}, \cdots$ 中可望会有非平稳性，因此简化算子 $\nabla_s z_t = (1-B^s)z_t = z_t - z_{t-s}$ 可能是有用的。稳定的非平稳算子 $1-B^s$ 有 s 个零点 $e^{i(2\pi k/s)}(k=0,1,\cdots,s-1)$，均匀地分布在单位圆上。另外，最终预报函数满足 $(1-B^s)\hat{z}_t(l)=0$，因而可以(但不是必须) 用完全互补的正弦和余弦来表示：

$$\hat{z}_t(l) = b_0^{(t)} + \sum_{j=1}^{[s/2]} \left[b_{1j}^{(t)} \cos\left(\frac{2\pi jl}{s}\right) + b_{2j}^{(t)} \sin\left(\frac{2\pi jl}{s}\right) \right]$$

这里 b 是适应系数，若 s 为偶数， $[s/2] = \frac{1}{2}s$；若 s 为奇数，$[s/2] = \frac{1}{2}(s-1)$.

乘积模型　假设有一个显示出季节特性且周期为已知的序列，我们把数据安排成含 s 列的表格形式，这样做是很有意义的，如表 9.1 所示，该表中所列是取对数后的航空线旅客数据。对于季节数据来说挑选适当的变换需要特别小心。在这个例子中(见 9.3.5 节)，对于数据的分析支持使用对数变换。表 9.1 的安排是为了强调这样的事实：对于周期性数据重要的时间间隔不是一个，而是两个。对这个例子来说，这两个时间间隔对应于月和年。具体说来，对于： (a) 在特定年份中连续月份的观测值， (b) 在连续年份中同一月份的观测值，我们希望在这些数据之间都有联系发生。这种情形有些象双重方差分析模型，可以料想同一列的观测和同一行的观测都有相似性。

参考表 9.1 中的航空旅客数据，季节效应意味着某一特定月份的观测值，例如 4 月，与以往 4 月份的观测值是相关的。假设第 t 个观测值 z_t 就是 4 月份的值，我们可以通过如下形式的模型将该观测值与以往 4 月份的观测值联系起来：

$$\Phi(B^s)\nabla_s^D z_t = \Theta(B^s)\alpha_t \tag{9.1.4}$$

表 9.1　国际航空旅客月度总数 (按千人计) 的自然对数 (序列 G)

	1 月	2 月	3 月	4 月	5 月	6 月
1949	4.718	4.771	4.883	4.860	4.796	4.905
1950	4.745	4.836	4.949	4.905	4.828	5.004
1951	4.977	5.011	5.182	5.094	5.147	5.182
1952	5.142	5.193	5.263	5.199	5.209	5.384
1953	5.278	5.278	5.464	5.460	5.434	5.493
1954	5.318	5.236	5.460	5.425	5.455	5.576
1955	5.489	5.451	5.587	5.595	5.598	5.753
1956	5.649	5.624	5.759	5.746	5.762	5.924
1957	5.753	5.707	5.875	5.852	5.872	6.045
1958	5.829	5.762	5.892	5.852	5.894	6.075
1959	5.886	5.835	6.006	5.981	6.040	6.157
1960	6.033	5.969	6.038	6.133	6.157	6.282

	7 月	8 月	9 月	10 月	11 月	12 月
1949	4.997	4.997	4.913	4.779	4.644	4.771
1950	5.136	5.136	5.063	4.890	4.736	4.942
1951	5.293	5.293	5.215	5.088	4.984	5.112
1952	5.438	5.489	5.342	5.252	5.147	5.268
1953	5.576	5.606	5.468	5.352	5.193	5.303
1954	5.710	5.680	5.557	5.434	5.313	5.434
1955	5.897	5.849	5.743	5.613	5.468	5.628
1956	6.023	6.004	5.872	5.724	5.602	5.724
1957	6.142	6.146	6.001	5.849	5.720	5.817
1958	6.196	6.225	6.001	5.883	5.737	5.820
1959	6.306	6.326	6.138	6.009	5.892	6.004
1960	6.433	6.407	6.230	6.133	5.966	6.068

这里 $s=12, \nabla_s = 1-B^s, \Phi(B^s)$ 和 $\Theta(B^s)$ 分别是 B^s 的 P 次和 Q 次多项式且满足平稳和可逆性条件。与此类似，模型

$$\Phi(B^s)\nabla_s^D z_{t-1} = \Theta(B^s)\alpha_{t-1} \tag{9.1.5}$$

可以用来把当前 3 月份的特征和以往 3 月份的观测值相联系，如此等等，对于 12 个月中的每一个月都可以这样做。并且，假设在这些月度模型中所含参数 Φ 和 Θ 对每个月都近似相等，这种假设也是合理的。

现在这些模型中的误差分量 $\alpha_t, \alpha_{t-1}, \cdots$ 通常不是不相关的。例如，1960 年 4 月的旅客总数和以往 4 月的旅客总数相关，与此同时它也会和 1960 年 3 月，2 月 1 月等等的总数相关。由此我们可以料想，(9.1.4) 中的 α_t 与 (9.1.5) 中的 α_{t-1} 相关，还与 α_{t-2} 相关，等等。因此，为了照顾到这种联系，我们引入第二个模型

$$\phi(B)\nabla^d \alpha_t = \theta(B) a_t \tag{9.1.6}$$

其中，a_t 是白噪声过程，$\phi(B)$ 和 $\theta(B)$ 分别是 B 的 p 次和 q 次多项式，且满足平稳和可逆性条件，并有 $\nabla = \nabla_1 = 1 - B$ 。

将 (9.1.6) 代入 (9.1.4)，我们最终得到一般的乘积模型

$$\phi_p(B)\Phi_P(B^s)\nabla^s \nabla_d^D z_t = \theta_q(B)\Theta_Q(B^s) a_t \tag{9.1.7}$$

这里，对于这个具体例子，$s = 12$ 。另外，在 (9.1.7) 中加上了下标 p, q, P, Q，是为了提醒读者各种算子的阶数。可以称所得到的乘积模型的阶为 $(p, d, q) \times (P, D, Q)$ 。为了考虑多重季节性，我们可以用类似的讨论去得到有三个或多个周期分量的模型。

9.2 用乘积 $(0,1,1)\times(0,1,1)_{12}$ 模型对航空旅客数据的描述

在本章的余下部分我们考虑前面引入的季节模型的基本形式，以及该类模型的预报能力。我们还要考虑将这种模型与数据相联系而引出的识别、估计和诊断检验问题。做这些工作并不需要新的理论，只不过是我们在第 6 章到第 8 章中详细讨论过的方法和思想的应用。在 9.2 节中我们从相当仔细地讨论一个具体例子着手。在 9.3 节中我们讨论对于一般季节模型特别需要注意的方面。

本节所要详细说明的是将 $(0,1,1) \times (0,1,1)$ 模型和序列 G 的航空旅客数据联系在一起。在 9.2.2 节中我们考虑模型本身；9.2.3 节讨论模型的识别；9.2.4 节讨论模型的拟合，最后在 9.2.5 节讨论模型的诊断检验。

9.2.1 乘积 $(0,1,1)\times(0,1,1)_{12}$ 模型

我们看到，对于一个非平稳而又不含季节分量的时间序列进行分析，能广泛适用且简单的随机模型就是 IMA(0,1,1) 过程。按照 9.1.3 节中的讨论，我们采用如下模型

$$\nabla_{12} z_t = (1-\Theta B^{12})\alpha_t$$

以便联系间隔为 1 年的 z 。进而，我们使用相似的模型

$$\nabla \alpha_t = (1-\theta B)\alpha_t$$

把相隔 1 个月的 α 联系起来，这里， θ 和 Θ 通常有不同的值。将这两个表达式合并，我们就得到乘积季节模型

$$\nabla\nabla_{12} z_t = (1-\theta B)(1-\Theta B^{12})a_t \qquad (9.2.1)$$

模型为 $(0,1,1)\times(0,1,1)_{12}$ 阶。该模型显然可以写为

$$z_t - z_{t-1} - z_{t-12} + z_{t-13} = a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13} \quad (9.2.2)$$

这个模型的可逆域由 $(1-\theta B)(1-\Theta B^{12}) = 0$ 的根在单位圆外这一条件所要求，它由如下不等式定义:

$$-1 < \theta < 1 \quad , \quad -1 < \Theta < 1$$

注意，在 (9.2.1) 右边滑动平均算子 $(1-\theta B)(1-\Theta B) = 1-\theta B-\Theta B^{12}+\theta\Theta B^{13}$ 的阶为 $q+sQ = 1+12(1) = 13$ 。

我们将在 9.2.3,9.2.4 及 9.2.5 节说明，取对数后的航空旅客数据可由一个形如 (9.2.1) 的模型很好地拟合，在充分的近似下有: $\hat{\theta} = 0.4, \hat{\Theta} = 0.6$, 以及 $\hat{\sigma}_a^2 = 1.34\times10^{-3}$. 以此作为前提，我们就可以考虑如何使用带有参数估计值的模型，并可以方便地对序列的未来值作预报。

9.2.2 预报

在第 4 章中我们看到，考虑一般的模型有三种不同的基本方式，每一种都在第 5 章中引出了考察预报的不同形式。我们现在就对季节模型 (9.2.1) 的预报考虑这三种方法。

差分方程方法　直接由差分方程本身来计算预报值是最好的办法。因此，由于

$$z_{t+l}=z_{t+l-1}+z_{t+l-12}-z_{t+l-13}+a_{t+l}-\theta a_{t+l-1}-\Theta a_{t+l-12}+\theta\Theta a_{t+l-13} \tag{9.2.3}$$

在令 $\theta=0.4,\Theta=0.6$ 后，原点 t 提前 l 期最小均方误差预报立刻给出为

$$\begin{aligned}\hat{z}_t(l)=[&z_{t+l-1}+z_{t+l-12}-z_{t+l-13}\\&+a_{t+l}-0.4a_{t+l-1}-0.6a_{t+l-12}+0.24a_{t+l-13}]\end{aligned} \tag{9.2.4}$$

正如第 5 章中一样，我们称

$$[z_{t+l}]=E[z_{t+l}|\theta,\Theta,z_t,z_{t-1},\cdots]$$

为 z_{t+l} 在原点 t 所取的条件期望。在上面表达式中，假设参数确切地已知，并假设序列 $z_t,z_{t-1},\cdots$ 的信息一直伸延到遥远的过去。

实际应用依赖于下述事实：

1. 对实际数据拟合所得的可逆模型，通常只能得出仅仅明显依赖于序列近期值的预报。

2. 由于估计误差而引起参数值的微小变化，预报值对此是不灵敏的。

现在有

$$[z_{t+j}]=\begin{cases}z_{t+j} & ,\quad j\leqslant 0\\ \hat{z}_t(j) & ,\quad j>0\end{cases} \tag{9.2.5}$$

$$[a_{t+j}]=\begin{cases}a_{t+j} & ,\quad j\leqslant 0\\ 0 & ,\quad j>0\end{cases} \tag{9.2.6}$$

因此，正如在第 5 章中一样，为了得到预报值，我们简单地用预报值来代替未知的 z，而用 0 来代替未知的 a 。已知的 a 当然是已计算出的提前 1 期外推预报误差，即 $a_t = z_t - \hat{z}_{t-1}(1)$ 。

例如，为了得到提前三个月的预报值，我们有

$$z_{t+3} = z_{t+2} + z_{t-9} - z_{t-10} + a_{t+3} - 0.4a_{t+2} - 0.6a_{t-9} + 0.24a_{t-10}$$

在原点 t 取条件期望:

$$\hat{z}_t(3) = \hat{z}_t(2) + z_{t-9} - z_{t-10} - 0.6a_{t-9} + 0.24a_{t-10}$$

即为

$$\hat{z}_t(3)=\hat{z}_t(2)+z_{t-9}-z_{t-10}-0.6[z_{t-9}-\hat{z}_{t-10}(1)]+0.24[z_{t-10}-\hat{z}_{t-11}(1)]$$

因此

$$\hat{z}_t(3) = \hat{z}_t(2) + 0.4z_{t-9} - 0.76z_{t-10} + 0.6\hat{z}_{t-10}(1) - 0.24\hat{z}_{t-11}(1) \tag{9.2.7}$$

这是用既往的 z 和既往 z 的预报来表示当前的预报值。虽然对每一预报提前期各自的表达式可以很容易地写出，但预报值的计算最好用单一的表达式 (9.2.4) 直接进行，该式右边的元素由 (9.2.5) 和 (9.2.6) 定义。

图 9.2 给出了提前期直到 36 个月的预报值，这些预报都是对任意选定的原点 1957 年 7 月而做的。我们看到，这一仅包含两个参数的简单模型忠实地再现了季节模式，并给出了极好的预报值。当然我们还记得，同所有从一般线性随机模型得到的预报值一样，这种预报函数是自适应的。当季节模式发生变化时，这种变化会适当地反映到预报值中去。我们注意到，当提前 1 个月的预报值太大时，则从这点起所有未来的预报值都有过大的趋势。这是可以想象的，因为正如附录 A5.1 中曾指出，从同一原点开始但不同提前期的预报误差之间是高度相关的。当然提前期很长的预报，例如 36 个月，必定包含相当大的误差。但是在实际中，一

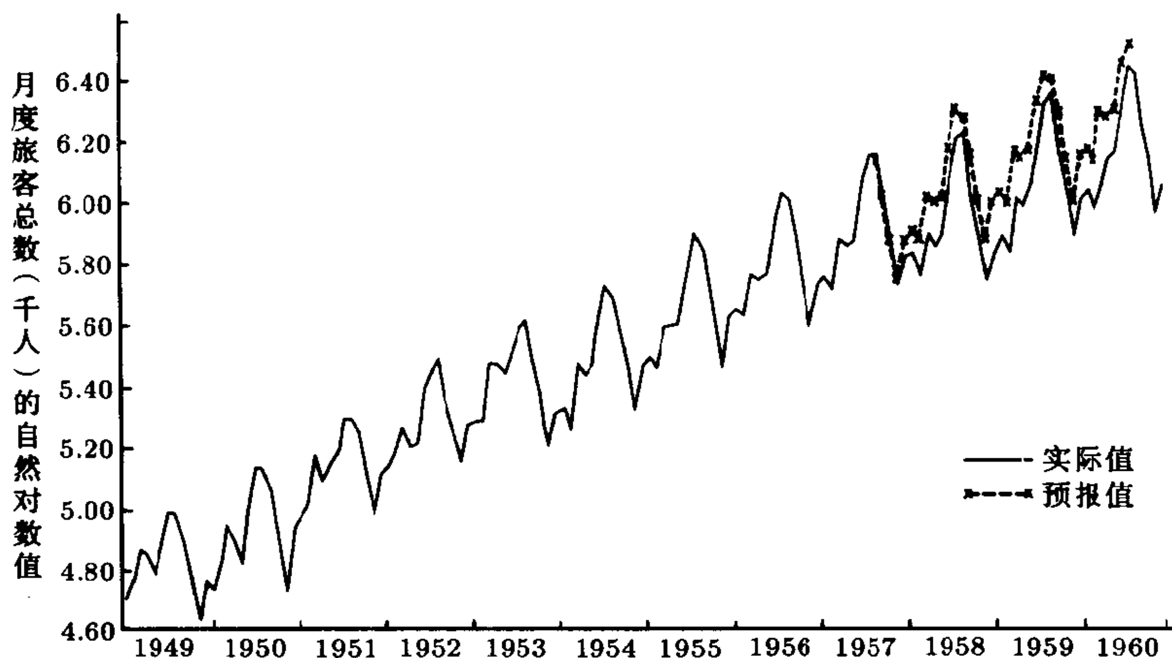


图 9.2 序列 G 的提前 1,2,3,⋯,36 个月的预报值，所有预报都是以任意选定的 1957 年 7 月为原点

开始是很遥远的预报值将会不断地被修正，并且随着预报提前期的缩短，可能会得到较好的精度。

上述预报方法对于参数值的适度变化是稳健的。因此，我们用 $\theta = 0.5, \Theta = 0.5$ 代替 $\theta = 0.4, \Theta = 0.6$，预报值不会受到很大的影响，即使对于提前多步的预报 (如 12 个月) 也是如此。通过研究平方和曲面可以看到调整参数值对提前一步预报的近似影响，因此可知 k 个参数 $\boldsymbol{\beta}$ 的近似置信域通常是由等值线 $S(\boldsymbol{\beta}) = S(\hat{\boldsymbol{\beta}})[1 + \chi^2_\epsilon(k)/n]$ 所界定的，它以概率 $1-\epsilon$ 包含真实参数。因此，我们知道，如果使用了参数真值，则具有同样概率的提前 1 期预报误差的均方不会增加到超过原来的 $1 + \chi^2_\epsilon(k)/n$ 倍。

预报函数及其修正，预报误差方差　正如我们在第 5 章中所述，在实际中，差分方程方法对于实际计算和修正预报值是最简单且最方便的。但是，有关由计算得到预报值的性质以及预报值的修正，差分方程方法本身并没有揭示出更多的东西。我们现在从其它的观点来考虑预报是为了阐明有关方面的问题，而不是给

出其它的计算方法.

预报函数 利用 (5.1.12) 得到

$$z_{t+l} = \hat{z}_t(l) + e_t(l) \tag{9.2.8}$$

其中

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{t-1} a_{t+1}$$

现在, (9.2.1) 右边的滑动平均算子的阶为 13, 因此, 由 (5.3.2) 和对于 $l > 13$, 预报值应满足差分方程

$$(1-B)(1-B^{12})\hat{z}_t(l) = 0 \qquad (l > 13) \tag{9.2.9}$$

在此方程中, B 是作用在 l 上.

我们现在记 $l = (r, m) = 12r + m, r = 0, 1, 2, \cdots$, 以及 $m = 1, 2, \cdots, 12$, 以此来表示预报提前期为 r 年零 m 个月, 例如, $l = 15 = (1, 3)$ 。于是预报函数是 (9.2.9) 的解, 该方程的初始条件由前 13 个预报值给出, 预报函数的形式为

$$\hat{z}_t(l) = \hat{z}_t(r, m) = b_{0,m}^{(t)} + r b_1^{(t)} \qquad (l > 0) \tag{9.2.10}$$

它包含 13 个可调系数 $b_{0,1}^{(t)}, b_{0,2}^{(t)}, \cdots, b_{0,12}^{(t)}, b_1^{(t)}$, 表示 12 个月度贡献值和 1 个年度贡献值, 它们是由前 13 个预报值确定. 由图 9.3 可以更清楚地了解预报函数的性质, 在图中给出了一个这类的函数, 只是周期为 $s = 5$, 因而有 6 个可调系数 $b_{0,1}^{(t)}, b_{0,2}^{(t)}, \cdots, b_{0,5}^{(t)}, b_1^{(t)}$ 。

等价地, 由于 $\hat{z}_t(l)$ 满足 (9.2.9), 且 $(1-B)(1-B^{12}) = 0$ 的根 $1, 1, -1, e^{\pm(i2\pi j/12)}, j = 1, 2, \cdots, 5$ 都在单位圆上, 如 (5.3.3) 中的预报函数能够表示为

$$\hat{z}_t(l) = \sum_{j=1}^{5}\left[b_{1j}^{(t)} \cos\left(\frac{2\pi j l}{12}\right) + b_{2j}^{(t)} \sin\left(\frac{2\pi j l}{12}\right)\right] + b_{16}^{(t)}(-1)^l + b_0^{(t)} + b_1^{*(t)} l$$

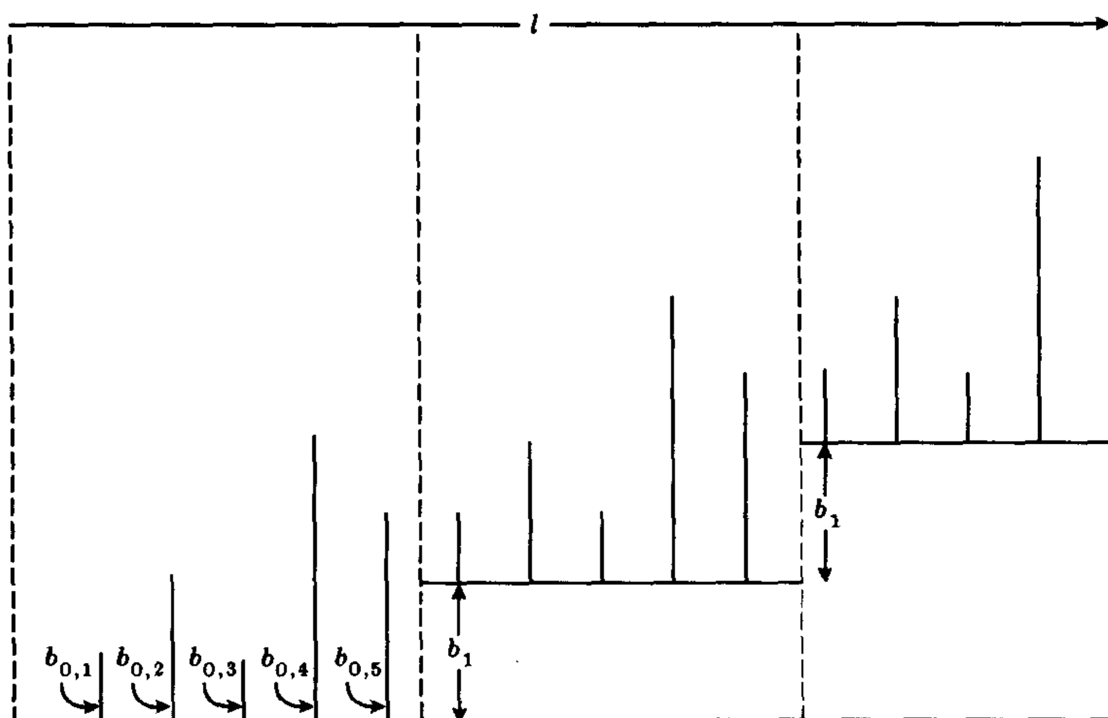


图 9.3　由模型 $\nabla\nabla_s z_t=(1-\theta B)(1-\Theta B^s)a_t$ 生成的季节性预报函数，其中 $S=5$

这表明 $\hat{z}_t(l)$ 是由季节频率为 $2\pi j/12, j=1,\cdots,6$ 的正弦曲线混合生成，另外还加上一个斜率为 $b_1^{*(t)}$ 的线性趋势。上述表达式中的系数 $b_{1j}^{(t)},b_{2j}^{(t)},b_0^{(t)}$ 和 $b_1^{*(t)}$ 关于预报原点 t 是自适应的，且由前 13 个预报值所确定。与 (9.2.10) 相比显然可见，例如，$b_1^{(t)}=12b_1^{*(t)}$，它表示在 $\hat{z}_t(l)$ 中的年度变化率，而 $b_1^{*(t)}$ 是月度变化率。

ψ 权　为了确定修正公式并得出预报误差 $e_t(l)$ 的方差，我们需要模型形式 $z_t=\sum_{j=0}^{\infty}\psi_j a_{t-j}$ 中的 ψ 权。我们可以将 (9.2.1) 中的滑动平均算子写成如下形式

$$(1-\theta B)(1-\Theta B^{12})=(\nabla+\lambda B)(\nabla_{12}+\Lambda B^{12})$$

这里，$\lambda=1-\theta,\Lambda=1-\Theta$。因此模型 (9.2.1) 可记为

$$\nabla\nabla_{12}z_t=(\nabla+\lambda B)(\nabla_{12}+\Lambda B^{12})a_t$$

通过在 $\nabla\nabla_{12}\psi(B)=(\nabla+\lambda B)(\nabla_{12}+\Lambda B^{12})$ 中令系数相等，可以看到 ψ 权满足：$\psi_0=1, \psi_1-\psi_0=\lambda-1, \psi_{12}-\psi_{11}-\psi_0=\Lambda-1, \psi_{13}-\psi_{12}-\psi_1+\psi_0=(1-\lambda)(1-\Lambda)$，以及 $\psi_j-\psi_{j-1}-\psi_{j-12}+\psi_{j-13}=0$ 等等. 因此对于该过程权 ψ 为

$$
\begin{array}{ll}
\psi_1=\psi_2=\cdots=\psi_{11}=\lambda & , \psi_{12}=\lambda+\Lambda \\
\psi_{13}=\psi_{14}=\cdots=\psi_{23}=\lambda(1+\Lambda) & , \psi_{24}=\lambda(1+\Lambda)+\Lambda \\
\psi_{25}=\psi_{26}=\cdots=\psi_{35}=\lambda(1+2\Lambda) & , \psi_{36}=\lambda(1+2\Lambda)+\Lambda
\end{array}
$$

等等。将 ψ_j 记为 $\psi_{r,m}=\psi_{12r+m}$，其中 $r=0,1,2,\cdots$ 及 $m=1,2,\cdots,12$ 分别指的是年和月，我们得到

$$\psi_{r,m}=\lambda(1+r\Lambda)+\delta\Lambda \tag{9.2.11}$$

其中

$$\delta=\begin{cases}1 & ,\text{当}m=12\\ 0 & ,\text{当}m\neq 12\end{cases}$$

修正　一般修正公式 (5.2.5) 为

$$\hat{z}_{t+1}(l)=\hat{z}_t(l+1)+\psi a_{t+1}$$

因此，如果 $m\neq s=12$, 则

$$b_{0,m}^{(t+1)}+rb_1^{(t+1)}=b_{0,m+1}^{(t)}+rb_1^{(t)}+(\lambda+r\lambda\Lambda)a_{t+1}$$

令 r 的系数相等，则修正公式为

$$
\begin{aligned}
b_{0,m}^{(t+1)}&=b_{0,m+1}^{(t)}+\lambda a_{t+1}\\
b_1^{(t+1)}&=b_1^{(t)}+\lambda\Lambda a_{t+1}
\end{aligned} \tag{9.2.12}
$$

另外，若 $m=s=12$, 有

$$b_{0,12}^{(t+1)}+rb_1^{(t+1)}=b_{0,1}^{(t)}+(r+1)b_1^{(t)}+(\lambda+\Lambda+r\lambda\Lambda)a_{t+1}$$

这时

$$
\begin{aligned}
b_{0,12}^{(t+1)} &= b_{0,1}^{(t)} + b_1^{(t)} + (\lambda + \Lambda)a_{t+1} \\
b_1^{(t+1)} &= b_1^{(t)} + \lambda\Lambda a_{t+1}
\end{aligned}
\tag{9.2.13}
$$

在考察这些关系时应该记住，$b_{0,m}^{(t+1)}$ 是 $b_{0,m+1}^{(t)}$ 的更新形式。因此如果原点 t 为某个特定年份的 1 月，那么 $b_{0,2}^{(t)}$ 就是 3 月的系数。当过了一个月后，我们应该把预报原点移到 2 月，从而 3 月的系数修正后为 $b_{0,1}^{(t+1)}$ 。

预报误差方差　权数 ψ 的信息使我们能够计算任何提前期的预报误差方差，利用 (5.1.16) 的结果，即是

$$
V(l) = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 \tag{9.2.14}
$$

在 (9.2.11) 和 (9.2.14) 中令 $\lambda = 0.6, \Lambda = 0.4, \sigma_a^2 = 1.34 \times 10^{-3}$ ，在表 9.2 中列出了航空旅客数据提前期为 1 到 36 的对数预报误差标准差估计值 $\hat{\sigma}(l)$ 。

预报作为既往观测值的加权平均　如果我们将模型写成如下

表 9.2　对于对数化航空旅客序列的各种提前期预报误差标准差的估计值

预报提前期	$\hat{\sigma}(l)\times10^{-1}$	预报提前期	$\hat{\sigma}(l)\times10^{-1}$	预报提前期	$\hat{\sigma}(l)\times10^{-1}$
1	3.7	13	9.0	25	14.4
2	4.3	14	9.5	26	15.0
3	4.8	15	10.0	27	15.5
4	5.3	16	10.5	28	16.0
5	5.8	17	10.9	29	16.4
6	6.2	18	11.4	30	17.0
7	6.6	19	11.7	31	17.4
8	6.9	20	12.1	32	17.8
9	7.2	21	12.6	33	18.3
10	7.6	22	13.0	34	18.7
11	8.0	23	13.3	35	19.2
12	8.2	24	13.6	36	19.6

形式

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t$$

提前 1 期预报误差是

$$\hat{z}_t(1) = \sum_{j=1}^{\infty} \pi_j z_{t+1-j}$$

π 权可通过在下式中令系数相等而得到:

$$(1 - B)(1 - B^{12}) = (1 - \theta B)(1 - \Theta B^{12})(1 - \pi_1 B - \pi_2 B^2 - \cdots)$$

于是

$$\begin{aligned}
&\pi_j = \theta^{j-1}(1 - \theta) \qquad (j = 1, 2, \cdots, 11) \\
&\pi_{12} = \theta^{11}(1 - \theta) + (1 - \Theta) \\
&\pi_{13} = \theta^{12}(1 - \theta) - (1 - \theta)(1 - \Theta) \\
&(1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})\pi_j = 0 \qquad (j > 14)
\end{aligned} \tag{9.2.15}$$

在图 9.4 中画出了这些权数，是关于参数 $\theta = 0.4$ 和 $\Theta = 0.6$ 给出的。

权函数取为图中所示具体形式，其原因可解释如下：过程 (9.2.1) 可以写成

$$a_{t+1} = \left(1 - \frac{\lambda B}{1 - \theta B}\right)\left(1 - \frac{\Lambda B^{12}}{1 - \Theta B^{12}}\right) z_{t+1} \tag{9.2.16}$$

我们现在用记号 $\text{EWMA}_\lambda(z_t)$ 来表示值 $z_t, z_{t-1}, z_{t-2}, \cdots$ 依参数 λ 的指数加权滑动平均，从而

$$\text{EWMA}_\lambda(z_t) = \frac{\lambda}{1 - \theta B} z_t = \lambda z_t + \lambda\theta z_{t-1} + \lambda\theta^2 z_{t-2} + \cdots$$

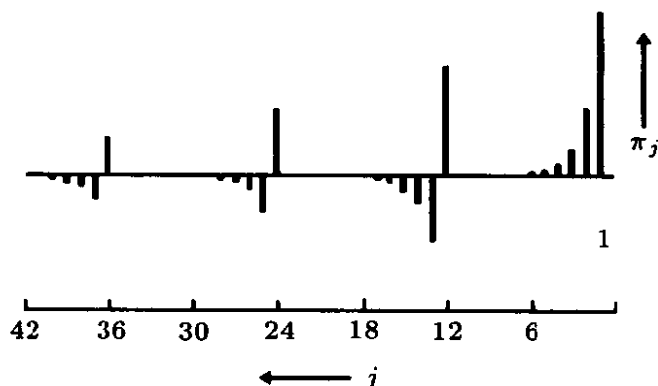


图 9.4 对序列 G 拟合 (0,1,1)×(0,1,1) 过程的 π 权 (θ=0.4,Θ=0.6)

类似地，我们用记号 $\mathrm{EWMA}_\Lambda(z_t)$ 来表示值 $z_t, z_{t-12}, z_{t-24}, \cdots$ 依参数 Λ 的指数加权滑动平均，从而有

$$\mathrm{EWMA}_\Lambda(z_t) = \frac{\Lambda}{1-\Theta B^{12}} z_t = \Lambda z_t + \Lambda\Theta z_{t-12} + \Lambda\Theta^2 z_{t-24} + \cdots$$

将 $\hat{z}_t(1) = z_{t+1} - a_{t+1}$ 代入 (9.2.16), 我们得到

$$\hat{z}_t(1) = \mathrm{EWMA}_\lambda(z_t) + \mathrm{EWMA}_\Lambda(z_{t-11} - \mathrm{EWMA}_\lambda z_{t-12}) \quad (9.2.17)$$

因此，预报值是对以往月份所取的 EWMA, 然后再用过去年份同样月度的 EWMA 与实际值之差的二次 EWMA 加以修正。作为 $\theta = 0(\lambda = 1)$ 的特殊情形， (9.2.17) 简化为

$$\begin{aligned}\hat{z}_t(1) &= z_t + \mathrm{EWMA}_\Lambda(z_{t-11} - z_{t-12}) \\ &= z_t + \Lambda[(z_{t-11} - z_{t-12}) + \Theta(z_{t-23} - z_{t-24}) + \cdots]\end{aligned}$$

该式表明，一阶差分的预报值就是对于过去年份同样月度一阶差分的季节 EWMA 。例如，假设我们要预测一个百货商店 12 月份的销售额。该销售额中包含了圣诞节购买这一重要成份。(9.2.17) 右边的第一项是对过去月份直到 11 月所取的 EWMA 。然而我们知道，这将是一个偏低的估计值，于是我们用下面方法加以修正：取过去年度 12 月份的实际销售额，再得出这些年中对过去月份所取相应的月度 EWMA, 用这二者之差对过去年份的二次 EWMA 作为修正。

用提前期较短的预报值来代替未知值，于是可由 π 权生成提前期 $l>1$ 的预报值，即

$$\hat{z}_t(2)=\pi_1\hat{z}_t(1)+\pi_2 z_t+\pi_3 z_{t-1}+\cdots$$

另外，直接加到 $z_t, z_{t-1}, z_{t-2}, \cdots$ 的权数的显式值也可以计算出，例如可由 (5.3.9) 或 (A5.2.3) 得到。

9.2.3 识别

非季节性 IMA(0,1,1) 过程的识别取决于这样的事实：在取一阶差分之后，所有超过 1 阶的自相关皆为零。对于乘积 $(0,1,1)\times(0,1,1)_{12}$ 过程来说，非零的自相关只是在滞后为 1,11,12 和 13 处。事实上，由 (9.2.2) 该模型可以看作是 $w_t=a_t-\theta a_{t-1}-\Theta a_{t-12}+\theta\Theta a_{t-13}$，这是一个 $w_t=\nabla\nabla_{12}z_t$ 的 13 阶 MA 模型。于是 w_t 的自协方差可由 3.3 节的结果直接得到，因此，它们是

$$\begin{aligned}
\gamma_0&=[1+\theta^2+\Theta^2+(\theta\Theta)^2]\sigma_a^2=(1+\theta^2)(1+\Theta^2)\sigma_a^2\\
\gamma_1&=[-\theta-\Theta(\theta\Theta)]\sigma_a^2=-\theta(1+\Theta^2)\sigma_a^2\\
\gamma_{11}&=\theta\Theta\sigma_a^2\\
\gamma_{12}&=[-\Theta-\theta(\theta\Theta)]\sigma_a^2=-\Theta(1+\theta^2)\sigma_a^2\\
\gamma_{13}&=\theta\Theta\sigma_a^2
\end{aligned}\tag{9.2.18}$$

特别，这些表达式蕴含 $\rho_1=-\theta/(1+\theta^2)$ 和 $\rho_{12}=-\Theta/(1+\Theta^2)$，因此，$\rho_1$ 值不受模型 (9.2.1) 中存在 MA 季节因子 $(1-\Theta B^{12})$ 的影响，而 ρ_{12} 值不受 (9.2.1) 中的非季节 (或正常) 因子 $(1-\theta B)$ 的影响。图 9.3 列出了取对数后航空旅客数据的自相关估计值，它们是对如下情形做出的：(a) 原始对数序列，(b) 仅按月差分的对数序列 ∇z_t，(c) 仅按年差分的对数序列 $\nabla_{12}z_t$,(d) 既按月也按年差分的对数序列 $\nabla\nabla_{12}z_t$。可以看到，z 的自相关数值大，而且对较高的滞后也不衰减。而简单的差分尽管一般会降低相关性，但仍保留非常强的周期成分。这尤其被滞后为 12,24 ,36 及 48 时非常大的相关性所证实。按周期为 12 的简单差分使得相关函

表 9.3 对数航空客票数据各种差分的自相关估计值

	滞后	自相关											
(a)z	1–12	0.95	0.90	0.85	0.81	0.78	0.76	0.74	0.73	0.73	0.74	0.76	0.76
	13–24	0.72	0.66	0.62	0.58	0.54	0.52	0.50	0.49	0.50	0.50	0.52	0.52
	25–36	0.48	0.44	0.40	0.36	0.34	0.31	0.30	0.29	0.30	0.30	0.31	0.32
	37–48	0.29	0.24	0.21	0.17	0.15	0.12	0.11	0.10	0.10	0.11	0.12	0.13
(b)∇z	1–12	0.20	–0.12	–0.15	–0.32	–0.08	0.03	–0.11	–0.34	–0.12	–0.11	0.21	0.84
	13–24	0.22	–0.14	–0.12	–0.28	–0.05	0.01	–0.11	–0.34	–0.11	–0.08	0.20	0.74
	25–36	0.20	–0.12	–0.10	–0.21	–0.06	0.02	–0.12	–0.29	–0.13	–0.04	0.15	0.66
	37–48	0.19	–0.13	–0.06	–0.16	–0.06	0.01	–0.11	–0.28	–0.11	–0.03	0.12	0.59
(c)$\nabla_{12} z$	1–12	0.71	0.62	0.48	0.44	0.39	0.32	0.24	0.19	0.15	–0.01	–0.12	–0.24
	13–24	–0.14	–0.14	–0.10	–0.15	–0.10	–0.11	–0.14	–0.16	–0.11	–0.08	0.00	–0.05
	25–36	–0.10	–0.09	–0.13	–0.15	–0.19	–0.20	–0.19	–0.15	–0.22	–0.23	–0.27	–0.22
	37–48	–0.18	–0.16	–0.14	–0.10	–0.05	0.02	0.04	0.10	0.15	0.22	0.29	0.30
(d)$\nabla\nabla_{12} z$	1–12	–0.34	0.11	–0.20	0.02	0.06	0.03	–0.06	0.00	0.18	–0.08	0.06	–0.39
	13–24	0.15	–0.06	0.15	–0.14	0.07	0.02	–0.01	–0.12	0.04	–0.09	0.22	–0.02
	25–36	–0.10	0.05	–0.03	0.05	–0.02	–0.05	–0.05	0.20	–0.12	0.08	–0.15	–0.01
	37–48	0.05	0.03	–0.02	–0.03	–0.07	0.10	–0.09	0.03	–0.04	–0.04	0.11	–0.05

数先是持续为正，而后又持续为负。对比之下，差分 $\nabla\nabla_{12}z_t$ 显著地降低了整个的相关性。

可以看到， $\nabla\nabla_{12}z_t$ 的自相关在滞后 1 和 12 时出现峰值，这与对应于模型 (9.2.1) 的 (9.2.18) 中的性质是一致的。但是，$\nabla_{12}z_t$ 的自相关在大于 1 的较低滞后可以看作是 (缓慢的) 指数率衰减，所以很可能 $\nabla_{12}z_t$ 并不是如 (9.2.1) 的非平稳 IMA(1,1) 模型，而是服从一个具有非季节算子形式如 ARMA(1,1) 的模型，其 ϕ 值相对接近于 1. 正如在第 6 章中所讨论的，在实际中这两种可能模型之间的差别是不显著的。在这里，我们对后一种可能模型不作进一步探讨。在非平稳和平稳的 AR(1) 因子之间做挑选，事实上可以通过类似于 6.3.6 节所述的方法来检验。

假设模型形式为 (9.2.1), 较高滞后自相关的方差由 Bartlett 公式 (2.1.13) 近似，在本例中为

$$\text{var}[r_k] \simeq \frac{1+2(\rho_1^2+\rho_{11}^2+\rho_{12}^2+\rho_{13}^2)}{n} \qquad (k>13) \qquad (9.2.19)$$

用估计相关来代替 ρ, 且在 (9.2.19) 中令 $n=144-13=131$, 这里 $n=131$ 是差分值 $\nabla\nabla_{12}z_t$ 的个数. 我们得到标准差 $\hat{\sigma}(r)\simeq 0.11$ 。

在表 9.4 中，列出了 35 个自相关 $r_k, k>13$ 的观察频数，并与具有零均值和标准差 0.11 的正态分布做比较. 粗略的检验表明对于该模型有必要做进一步的考察。

初估计 与处理非季节模型一样，令观察的相关函数等于它们的期望值，可以得到参数 θ 和 Θ 的近似值. 将样本估计值

表 9.4 对于 $\nabla\nabla_{12}z_t$ 滞后大于 13 的自相关，观察频数和期望频数的对比

	零均值标准差 0.11 的正态分布预期值	观测值 *
$0<\|r\|<0.11$	23.9	27.5
$0.11<\|r\|<0.22$	9.5	7.0
$0.22<\|r\|$	1.6	0.5
	35.0	35.0

* 对分区边界上的观察值将其各分一半到相邻的分区上.

$r_1 = -0.34$ 和 $r_{12} = -0.39$ 代入下面表达式

$$\rho_1 = \frac{-\theta}{1+\theta^2} \quad , \quad \rho_{12} = \frac{-\Theta}{1+\Theta^2}$$

我们得到粗估计 $\hat{\theta} \simeq 0.39$ 和 $\hat{\Theta} \simeq 0.48$ 。在附录 A9.1 给出了一个表，对于一些典型季节模型总结了自相关函数的性质，这对于识别和得到参数的初估计是有用的。

9.2.4 估计

图 9.5 给出了对航空旅客数据拟合模型 (9.2.1) 的平方和函数等值线，同时给出对应的 95% 置信域。可以看到，最小二乘估计非常接近于 $\hat{\theta} = 0.4$ 和 $\hat{\Theta} = 0.6$ 。$S(\theta, \Theta)$ 的格点值是用第 7 章所述方法计算的。在那里说明了，若 n 个观测值 w 是由下面线性过程所定义：

$$\phi(B)w_t = \theta(B)a_t$$

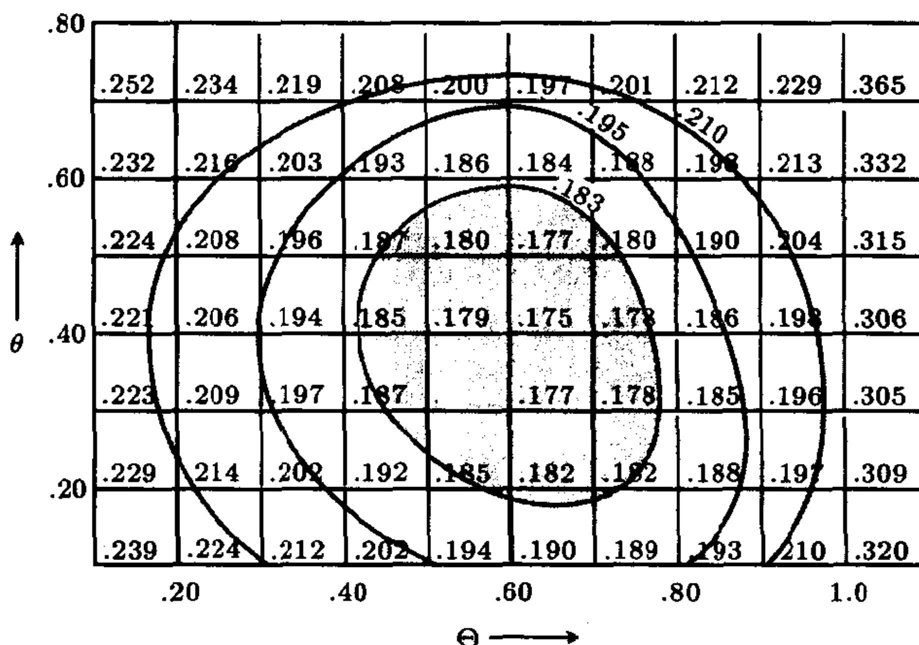


图 9.5 用模型 $\nabla\nabla_{12}z_t=(1-\theta B)(1-\Theta B)a_t$ 拟合序列 G:$S(\theta,\Theta)$ 的等值线及阴影所示的 95% 置信域

表 9.5 航空旅客数据：对 $[a]$ 的 $S(\theta,\Theta)$ 的计算表

z_t	t	$[a_t]$	$[w_t]$	$[e_t]$	a_t^0	u_t
		0	0	0		
z_{-12}	-12	$[a_{-12}]$	$[w_{-12}]$	0	0	
z_{-11}	-11	$[a_{-11}]$	$[w_{-11}]$	0	0	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
z_0	0	$[a_0]$	$[w_0]$	0	0	
z_1	1	$[a_1]$	$[w_1]$	$[e_1]$	a_1^0	u_1
z_2	2	$[a_2]$	$[w_2]$	$[e_2]$	a_2^0	u_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
z_{131}	131	$[a_{131}]$	$[w_{131}]$	$[e_{131}]$	a_{131}^0	u_{131}

那么，出现在似然函数中指数部分的二次型 $\boldsymbol{w}'\boldsymbol{M}_n\boldsymbol{w}$ 总可以表示成两部分之和： a 的条件期望平方和加上 $p+q$ 个初值 $\boldsymbol{e}_* = (w_{t-p},\cdots,w_0,a_{1-q},\cdots,a_0)'$ 条件期望的二次函数。即

$$\boldsymbol{w}'\boldsymbol{M}_n\boldsymbol{w} = S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=-\infty}^{n}[a_t]^2 = \sum_{r=1}^{n}[a_t]^2 + [\boldsymbol{e}_*]'\boldsymbol{\Omega}^{-1}[\boldsymbol{e}_*]$$

其中 $[a_t]=[a_t|\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}]$ 及 $[\boldsymbol{e}_*]=[\boldsymbol{e}_*|\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\theta}]$ 。而且， $S(\boldsymbol{\phi},\boldsymbol{\theta})$ 在 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的参数估计中起着重要的作用，不论是从样本理论还是似然函数或 Bayes 观点来说明，都是如此。

对于季节模型的计算与 7.1.5 节中所述对非季节模型的方法是完全一样的。我们通过对航空旅客数据计算 $S(\theta,\Theta)$, 并考虑与下面模型相联系：

$$\nabla\nabla_{12}z_t = w_t = (1-\theta B)(1-\Theta B^{12})a_t$$

以此来加以说明。表 9.5 列出了一个计算 $[a_t]$ 的格式。如果有 N 个 z 的观测值，那么，通常由差分算子 $\nabla^d\nabla_s^D$ 我们可以算出 $n=N-d-sD$ 个 w 值。因此，采用一种标号系统使得序列 z 中的第一个观测具有下标 $1-d-sD$, 这样做是很方便的。于是序列 w 中的第一个观测具有下标 1, 而最后一个下标是 n 。

在航空旅客序列中有 $N=144$ 个观测值. 因此在表 9.5 中它们表示为 $z_{-12}, z_{-11}, \cdots, z_{131}$. 因而由差分得到的 w 表示为序列 $w_1, w_2, \cdots, w_n$, 这里 $n=131$. 这些值排列在表的中央.

正如 7.1.5 节所指出的, 为计算方便起见, 可以使用由相应的反向模型来求得 $[a_t]$ 的近似方法. 在这种方法中, 正向和反向递推的基本公式可以和前面一样获得, 即在模型的正向和反向形式上取条件期望而得到. 在这个例子中, 这些公式为

$$[e_t]=[w_t]+\theta[e_{t+1}]+\Theta[e_{t+12}]-\theta\Theta[e_{t+13}] \quad (9.2.20)$$
$$[a_t]=[w_t]+\theta[a_{t-1}]+\Theta[a_{t-12}]-\theta\Theta[a_{t-13}] \quad (9.2.21)$$

通常, 对于季节模型来说我们可以有 $(p+sP)$ 阶的平稳自回归算子. 如果我们希望 7.1.5 节中的反向计算在序列中尽可能靠后地开始, 那么, 迭代就可以由 $[e_{n-p-sP}]$ 的近似值开始计算, 其中设未知的 $[e]$ 等于零. 在本例中, $p=P=0$, 由此, 利用 (9.2.20), 我们可开始于:

$$[e_{131}]=w_{131}+(\theta\times 0)+(\Theta\times 0)-(\theta\Theta\times 0)$$
$$[e_{130}]=w_{130}+(\theta\times[e_{131}])+(\Theta\times 0)-(\theta\Theta\times 0)$$

依此类推, 直到求得 $[e_1]$. 注意, 当 $j>0$ 时 $[e_{-j}]=0$, 我们现在可由 (9.2.20) 计算反向预报值 $[w_0], [w_{-1}], \cdots, [w_{-12}]$. 此外, 对于 $j>12, [w_{-j}]$ 的值皆为零, 并且因为每个 $[a_t]$ 都是过去发生的 $[w]$ 的函数, 所以 $[a_{-j}]=0, j>12$(而其它值显然可以直接由模型的形式得出) . 因此, (9.2.21) 现在可以直接用来计算 $[a]$, 从而计算 $S(\theta,\Theta)=\sum_{t=-12}^{131}[a_t]^2$. 在几乎所有有意义的例子中, 由于反向递推开始时的近似将会导致瞬态变化, 但它对于初始 $[w]$ 计算的影响是可以忽略的, 因此由这种方法计算的 $S(\theta,\Theta)$ 基本上是准确的. 正如在 7.1.5 节中所指出的, 这样一来我们就可以持续不断地"向前、向后"迭代了. 在下一次迭代时将要涉及再次计算 $[e]$, 而开始迭代要用到预报值 $[w_{n+1}], [w_{n+2}], \cdots, [w_{n+13}]$, 这些是由已经算出的 $[a]$ 得到的.

另外，正如 7.1.5 节所讨论的，可以使用精确方法来方便地计算 $[a_t]$ 。对于当前的模型来说，这首先需要使用零初值 $a^0_{-12} = a^0_{-11} = \cdots = a^0_0 = 0$ 来计算 $[a_t]$ 的条件估计值，通过与 (9.2.21) 类似的递推公式，即

$$a^0_t = w_t + \theta a^0_{t-1} + \Theta a^0_{t-12} - \theta\Theta a^0_{t-13} \qquad (t = 1, \cdots, n)$$

为了由 a^0_t 得出 u_t 的值，可用向后递推公式

$$u^0_t = a^0_t + \theta u_{t+1} + \Theta u_{t+12} - \theta\Theta u_{t+13} \qquad (t = n, \cdots, 1)$$

其中使用零初值 $u_{n+1} = \cdots = u_{n+13} = 0$ 。最后，正如附录 A7.3 的 (A7.3.12) 所做的一般描述，初始向量 $\boldsymbol{a}'_* = (a_{-12}, \cdots, a_0)$ 的精确反向预报值通过求解方程组 $\boldsymbol{D}[\boldsymbol{a}_*] = \boldsymbol{F}'\boldsymbol{u}$ 得到。我们注意到，如果向量 $\boldsymbol{F}'\boldsymbol{u}$ 表示为 $\boldsymbol{h} = \boldsymbol{F}'\boldsymbol{u} = (h_{-12}, h_{-11}, \cdots, h_0)'$ ，则 h_{-j} 值计算为 $h_{-j} = -(\theta u_{-j+1} + \Theta u_{-j+12} - \theta\Theta u_{-j+13})$，这里必须用到约定 $u_{-j} = 0, j > 0$ 。一旦求得反向预报值，剩下的 $[a_t]$ 值 $t = 1, 2, \cdots, n$ 完全和前面的方法一样可由 (9.2.21) 递推得到，从而精确的平方和 $S(\theta, \Theta) = \sum_{t=-12}^{131} [a_t]^2$ 可以得到。由这种方法计算 $[a_t]$ 的格式在表 9.5 的最后两行给出。

最小二乘估计 $\hat{\theta}, \hat{\Theta}$ 的迭代计算　正如在 7.2 节中所讨论的，从本质上说，在新的情况下或是出现困难时，画出平方和曲面才是必不可少的，在一些简单明白的场合可以利用线性化的迭代方法来给出最小二乘估计及其近似标准差。这种方法在 7.2.1 节中已给出，在估计季节模型参数时不会出现新的困难。

对于当前的例子，我们可以近似地写为

$$a_{t,0} = (\theta - \theta_0)x_{t,1} + (\Theta - \Theta_0)x_{t,2} + a_t$$

这里

$$x_{t,1} = -\left.\frac{\partial a_t}{\partial \theta}\right|_{\theta_0, \Theta_0}, \quad x_{t,2} = -\left.\frac{\partial a_t}{\partial \Theta}\right|_{\theta_0, \Theta_0}$$

表 9.6　对于对数航空旅客数据 θ 和 Θ 的迭代估计

迭代次数	θ	Θ
初始值	0.390	0.480
1	0.404	0.640
2	0.395	0.612
3	0.396	0.614
4	0.396	0.614

其中，θ_0 和 Θ_0 是推测值，$a_{t,0}=[a_t|\theta_0,\Theta_0]$。如在 7.2.2 节中所解释和说明的，导数用数值计算是最容易的。利用 9.2.3 节中由自相关估计值得到初估计 $\hat{\theta}=0.39, \hat{\Theta}=0.48$ 作为初值，按上述方法着手进行，迭代过程如表 9.6 所示。另外，如在节 7.2.3 中所做的，通过递推计算可以得到任意要求精度的导数值。

这样，精确到两位小数的参数值可在三次迭代后获得，这种精度是在实际中需要最多的。残差方差的估计值为 $\hat{\sigma}_a^2=1.34\times 10^{-3}$。从最后一次迭代中平方和矩阵的逆乘以 x 可以计算出估计值的标准差。于是最小二乘估计及其标准差为

$$\hat{\theta}=0.4\pm 0.08$$
$$\hat{\Theta}=0.61\pm 0.07$$

这与平方和图上所得值相当吻合。

估计值的大样本方差和协方差　和 7.2.6 节中一样，可以得到参数估计值方差和协方差的大样本公式。在本例中，由模型方程 $w_t=a_t-\theta a_{t-1}-\Theta a_{t-12}+\theta\Theta a_{t-13}$ 可看出导数 $x_{t,1}=-\partial a_t/\partial\theta$ 满足

$$x_{t,1}-\theta x_{t-1,1}-\Theta x_{t-12,1}+\theta\Theta x_{t-13,1}+a_{t-1}-\Theta a_{t-13}=0$$

因此 $(1-\theta B)(1-\Theta B^{12})x_{t,1}=-(1-\Theta B^{12})a_{t-1}$，或简单地有 $(1-\theta B)x_{t,1}=-a_{t-1}$。于是对 $x_{t,2}=-\partial a_t/\partial\Theta$ 用类似的推导，我们

得到

$$x_{t,1} \simeq -(1-\theta B)^{-1}a_{t-1} = -\sum_{j=0}^{\infty}\theta^j B^j a_{t-1}$$

$$x_{t,2} \simeq -(1-\Theta B^{12})^{-1}a_{t-12} = -\sum_{i=0}^{\infty}\Theta^i B^{12i} a_{t-12}$$

故对于大样本情形，信息阵为

$$I(\theta,\Theta) = n\begin{bmatrix}(1-\theta^2)^{-1} & \theta^{11}(1-\theta^{12}\Theta)^{-1}\\ \theta^{11}(1-\theta^{12}\Theta)^{-1} & (1-\Theta^2)^{-1}\end{bmatrix}$$

假设 $|\theta|$ 接近于 1, 则非对角线项可以忽略，$\hat{\theta}$ 和 $\hat{\Theta}$ 的方差和协方差的近似值为

$$V(\hat{\theta}) \simeq n^{-1}(1-\theta^2) \quad , \quad V(\hat{\Theta}) \simeq n^{-1}(1-\Theta^2)$$
$$\operatorname{cov}[\hat{\theta},\hat{\Theta}] \simeq 0 \tag{9.2.22}$$

在当前例子中，用 $\hat{\theta}=0.4, \hat{\Theta}=0.61$ 以及 $n=131$ 代入，得到

$$V(\hat{\theta}) \simeq 0.0064 \quad , \quad V(\hat{\Theta}) \simeq 0.0048$$

及

$$\sigma(\hat{\theta}) \simeq 0.08 \quad , \quad \sigma(\hat{\Theta}) \simeq 0.07$$

在此精度下，它与直接由迭代所得的值相等。我们注意到，分别与月度和年度相联系的参数 $\hat{\theta}$ 和 $\hat{\Theta}$ 实际上是不相关的，这也是很有趣的。

9.2.5 诊断检验

在进行下一步之前，我们通过考察拟合过程的残差来检验模型拟合的恰当性。

自相关检验　表 9.7 给出了残差 $\hat{a}_t = \nabla\nabla_{12}z_t + 0.40\hat{a}_{t-1} + 0.61\hat{a}_{t-12} - 0.24\hat{a}_{t-13}$ 的自相关估计值。许多单个相关与它们的标

表 9.7 对于对数航空旅客数据 (序列 G) 拟合残差的自相关估计值，模型为 $\nabla\nabla_{12}z_t=(1-0.4B)(1-0.61B^{12})a_t$

滞后 k	自相关 $r_k(\hat{a})$						标准差 (上界)
1- 6	0.02	0.02	-0.13	-0.14	0.05	0.06	0.09
7-12	-0.07	-0.14	0.10	-0.08	0.02	-0.01	0.09
13-18	0.03	0.04	0.05	-0.16	0.03	0.00	0.09
19-24	-0.11	-0.10	-0.03	-0.03	0.22	0.03	0.09
25-30	-0.02	0.06	-0.04	-0.06	-0.05	-0.08	0.09
31-36	-0.05	0.12	-0.13	0.00	-0.06	-0.02	0.09
37-42	0.11	0.07	-0.02	-0.05	-0.10	-0.02	0.09
43-48	-0.04	0.00	-0.08	0.03	0.04	0.06	0.09

$$(n+2)\sum_{k=1}^{24} r_k^2(\hat{a})/(n-k)=0.1950$$

准差上界 0.09 相比显得较大，值 $r_{23}=0.22$ 尤其异常，它大约是上界的 2.5 倍。然而，在 48 个随机偏差中出现一些大的偏差也是预料之中的。

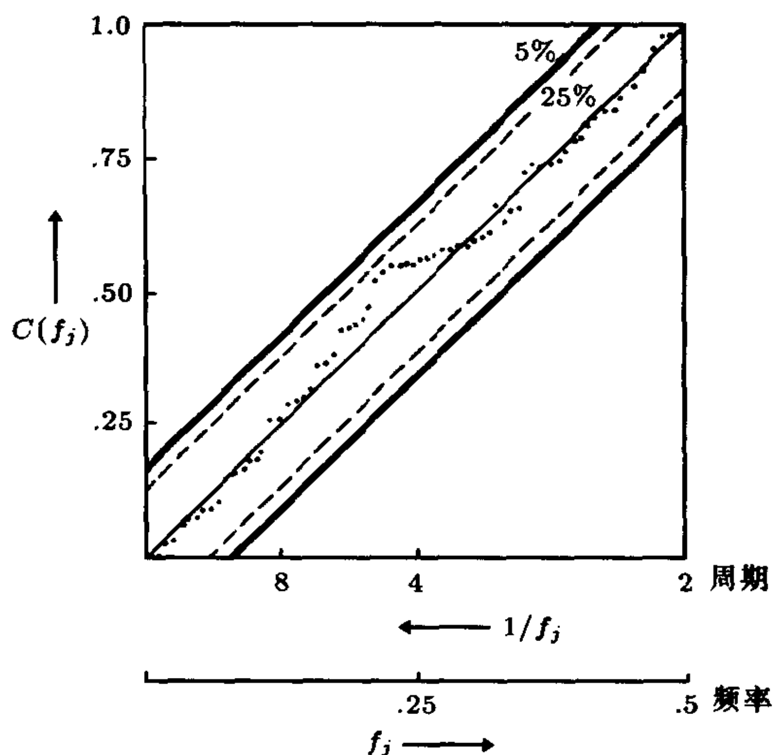


图 9.6 对于序列 G 拟合模型 $\nabla\nabla_{12}z_t=(1-0.4B)(1-0.61B^{12})a_t$ 所得残差的累积周期图检验

量 $\widetilde{Q}=n(n+2)\sum_{k=1}^{24}r_k^2(\hat{a})/(n-k)$ 给出了一个总体检验，由于拟合了两个参数，所以它近似于自由度为 22 的 χ^2 分布 (见 8.2.2 节)。 $\widetilde{Q}$ 的观察值为 $131\times 0.1950=25.5$, 在模型为恰当的假设之下，比这个偏差更大的情形可望为 27%, 因此，上述检验不能为模型不恰当提供任何依据。

周期图检验 图 9.6 给出了残差的累积周期图 (见 8.2.5 节)。正如我们在 8.2.5 节中所见到的， Kolmogorov-Smirnov 的 5% 和 25% 的概率限为明显偏差的限著性提供了一个粗略的指示，在本例中，它不能指出对假设模型的任何显著偏差。

9.3 更一般季节模型的某些方向

9.3.1 乘积和非乘积模型

在前几节中我们已经讨论了处理季节时间序列的方法，特别是以考察一个乘积模型作为例子。我们已经看到，如何用非常少的参数给出有效的表达式。现在我们尚须研究这一类的其它季节模型，以及随着新的考虑而产生的有关识别、估计、诊断检验及预测方法。

通常假设存在周期为 s 的季节影响。那么，乘积模型的一般形式可以用图 9.7 中给出的方式来说明。在乘积模型中，假设在“周期之间”序列的演化用如下模型表示

$$\Phi_P(B^s)\nabla_s^D z_{r,m}=\Theta_Q(B^s)\alpha_{r,m}$$

而在周期内的 α 通过下式联系

$$\phi_p(B)\nabla^d\alpha_{r,m}=\theta_q(B)\alpha_{r,m}$$

显然，我们可以改换所考虑的这两种模型的顺序，两种情形都得到一般的乘积模型

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_{r,m}=\theta_q(B)\Theta_Q(B^s)a_{r,m} \qquad (9.3.1)$$

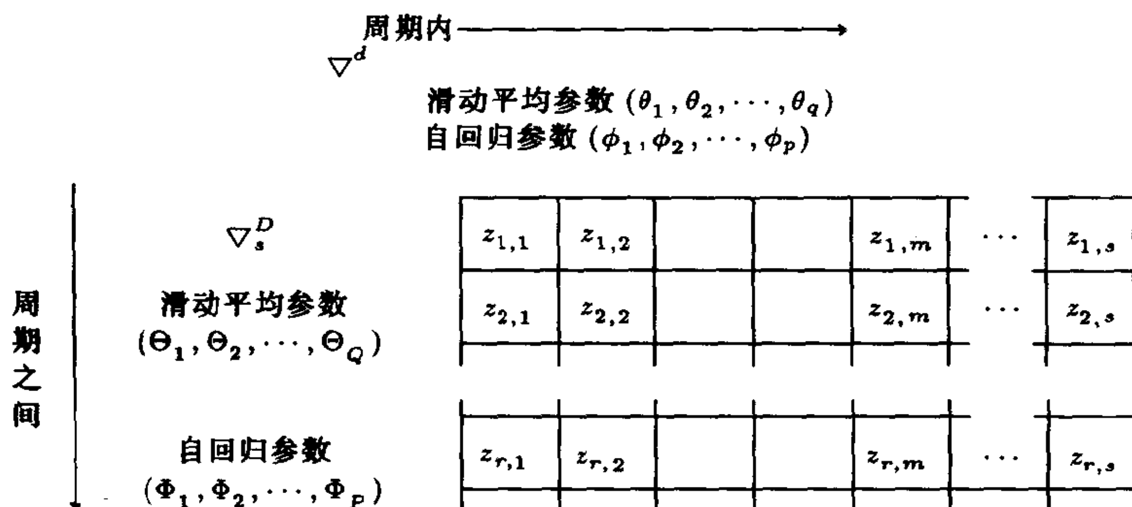


图 9.7 对于乘积季节模型两种处理方式的图表

这里 $a_{r,m}$ 是零均值白噪声过程. 在实际中，如 (9.3.1) 这类模型的有效性取决于在多大程度上可以对实际时间序列的一些项进行简约的参数化。事实上，对于来源广为不同的多种季节时间序列 ([18]) 这是可能的。

对于所有序列都用乘积模型来得到完全恰当的拟合是不可能的。这样的一种修正有时是有用的，即允许混合滑动平均算子为非乘积性的。这意味着我们在 (9.3.1) 的右边用一个更一般的滑动平均算子 $\theta^*_{q^*}(B)$ 来代替 $\theta_q(B)\Theta_Q(B)$ 。另外，在左边用一个更一般的算子 $\phi^*_{p^*}(B)$ 来代替 $\phi_p(B)\Phi_P(B^s)$ 也可能是必要的。在附录 A9.1 中给出了一些非乘积模型的典型形式，它们的标号是 4,4a,5 和 5a 。

在非乘积模型是必要的那些场合，经验表明：拟合一个好的乘积模型可以提供一个良好的开端，由此出发再构造更好的非乘积模型。这种情况使我们联想到分析两因素方差分析表时所遇到的问题，在那里，行和列常数的加性可能并不是一种恰当的假设，但可以给出一个很好的出发点。

关于对数据建立乘积或非乘积模型的一般策略，我们已在 9.2 节中作了较详细的讨论和说明。那就是用自相关函数作为先导：

1. 对于序列作差分 ∇ 和 ∇_s, 以得到平稳性。

2. 检查适当差分后序列的自相关函数, 选择一个尝试性的模型。

3. 由差分序列的适当自相关值求得参数的初估计。这些值可用于寻找最小二乘估计的初始值。

4. 经拟合后, 对残差运用诊断检验过程, 或者导致接受试验性模型, 或者指出可以改进的方式, 进而重新拟合及重复进行诊断检验。

作为对模型识别的一点实际指导, 我们注意到对于季节序列所需季节差分 D 通常不会超过 1 阶, 特别是对于 $s=12$ 的月度序列, 季节 AR 算子 $\Phi(B^s)$ 和 MA 算子 $\Theta(B^s)$ 的阶数很少超过 1 阶。当可利用数据的序列长度不足以支持 $P>1$ 或 $Q>1$ 的复杂模型形式时, 尤其如此。

9.3.2 识别

附录 A9.1 中的表在模型识别中应是一种有用的辅助手段, 表中对许多简单的季节模型给出了 $w_t=\nabla^d\nabla_S^D z_t$ 的协方差结构。该表并不是全能的, 但是, 可以相信它的确包含了经常遇到的模型, 并且读者会不难发现其它有代表意义的模型特征。应该强调的是, 一些相当简单的模型 (例如附录 A9.1 中的模型 1 和 2) 已对许多季节序列给出了恰当的表达式。

因为差分序列 $w_t=\nabla\nabla_s z_t$ 的乘积季节 ARMA 模型可以被看作是阶为 $p+sP$ 和 $q+sQ$ 的 ARMA 模型的特殊形式, 所以它们的自协方差函数可由第 3 章中的理论推出, 正如 9.2.3 节中对 MA 模型 $w_t=a_t-\theta a_{t-1}-\Theta a_{t-12}+\theta\Theta a_{t-13}$ 所作的一样。为了说明附录 A9.1 中所给出模型自协方差函数的推导, 考虑模型 $(1-\phi B)w_t=(1-\Theta B^s)a_t$, 它是 AR 阶为 1 、 MA 阶为 s 的特殊形式的 ARMA 模型。首先, 由于这个关于 w_t 模型的 ψ 权满足 $\psi_j-\phi\psi_{t-1}=0, j=1,\cdots,s-1$, 所以我们有 $\psi_j=\phi^j, j=$

$1, \cdots, s-1$, 并有 $\psi_s = \phi^s - \Theta$ 及 $\psi_j = \phi\psi_{j-1}, j > s$ 。由 3.4 节我们知道 w_t 的自协方差函数满足

$$\begin{aligned} &\gamma_0 = \phi\gamma_1 + \sigma_a^2(1 - \Theta\psi_s) \\ &\gamma_j = \phi\gamma_{j-1} - \sigma_a^2\Theta\psi_{s-j} \qquad (j = 1, \cdots, s) \\ &\gamma_j = \phi\gamma_{j-1} \qquad (j > s) \end{aligned} \tag{9.3.2}$$

对 γ_0 和 γ_1 求解前两个方程，我们得到

$$\gamma_0 = \sigma_a^2 \frac{1 - \Theta(\phi^s - \Theta) - \phi^s\Theta}{1 - \phi^2} = \sigma_a^2 \frac{1 + \Theta^2 - 2\phi^s\Theta}{1 - \phi^2}$$

$$\gamma_1 = \sigma_a^2 \frac{\phi[1 - \Theta(\phi^s - \Theta)] - \phi^{s-1}\Theta}{1 - \phi^2} = \sigma_a^2 \frac{\phi(1 + \Theta^2 - \phi^s\Theta) - \phi^{s-1}\Theta}{1 - \phi^2}$$

并有 $\gamma_j = \phi\gamma_{j-1} - \sigma_a^2\Theta\phi^{s-j} = \phi^j\gamma_0 - \sigma_a^2\Theta\phi^{s-j}(1+\phi^{2j})/(1-\phi^2), j = 1, \cdots, s$, 以及 $\gamma_j = \phi\gamma_{j-1} = \phi^{j-s}\gamma_s, j > s$ 。因此，特别对于 $s = 12$ 和 $|\phi|$ 不太接近 1 的月度数据来说，对于较低的滞后，该过程的自相关函数 ρ_j 特性类似于正规的 AR(1) 过程，即对于小的 j 有 $\rho_j \simeq \phi^j$, 但 ρ_{12} 的值接近于 $-\Theta/(1+\Theta^2)$ 。

在推导乘积过程自协方差时一个非常有用的事实是：对这种乘积过程来说，自协方差生成函数 (3.1.11) 是其分量生成函数的乘积。因此，在 (9.3.1) 中如果 $\nabla^d z_t$ 和 $\nabla_s^D \alpha_t$ 的分量模型

$$\phi_p(B)\nabla^d z_t = \theta_q(B)\alpha_t \ , \qquad \Phi_P(B^s)\nabla_s^D \alpha_t = \Theta_Q(B^s)a_t$$

的自协方差生成函数为 $\gamma(B)$ 和 $\Gamma(B)$, 则在 (9.3.1) 中 $w_t = \nabla^d\nabla_s^D z_t$ 的自协方差生成函数为

$$\gamma(B)\Gamma(B^s)$$

在模型 4 和 5 中使用的三个参数的非乘积算子为

$$1 - \theta_1 B - \theta_{12}B^{12} - \theta_{13}B^{13} \tag{9.3.3}$$

可以写成

$$(1-\theta_1 B)(1-\theta_{12}B^{12})-\kappa B^{13}$$

其中

$$\kappa=\theta_1\theta_{12}-(-\theta_{13})$$

κ 的估计值与其标准差相比较大时，说明需要一个非乘积模型，其中 θ_{13} 与 θ_1 和 θ_{12} 的联系不紧密。另一方面，如果 κ 较小，可记 $\theta_1=\theta,\theta_{12}=\Theta$, 则模型近似于乘积 $(0,1,1)\times(0,1,1)_{12}$ 模型。

9.3.3 估计

在一般季节模型的参数估计中不会出现新的问题。无条件平方和可用完全通用的方法进行计算，有关方法在 7.1.5 节中详细给出并在 9.2.4 节做了进一步的说明。如通常一样，绘出等值线图对存在困难的情形会有所启发。而在特性较好的场合，用数值导数的迭代最小二乘会很快地收敛到最小二乘估计值，并同时得到估计值的方差和协方差。如果需要的话，在任何情形都能推出可以直接计算导数的递推算法。

估计值的大样本方差和协方差　大样本信息阵 $I(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{\Phi},\boldsymbol{\Theta})$ 可通过计算 $E[\boldsymbol{X}'\boldsymbol{X}]$ 而给出，和 7.2.6 节中一样，在这里 $\boldsymbol{X}$ 是与导数阵符号相反的 $n\times(p+q+P+Q)$ 阵。那么，对于一般乘积模型

$$a_t=\theta^{-1}(B)\Theta^{-1}(B^s)\phi(B)\Phi(B^s)w_t$$

其中

$$w_t=\nabla^d\nabla_S^D z_t$$

所需要的导数为

$$\frac{\partial a_t}{\partial\theta_i}=\theta^{-1}(B)B^i a_t \quad , \quad \frac{\partial a_t}{\partial\Theta_i}=\Theta^{-1}(B^s)B^{si}a_t$$

$$\frac{\partial a_t}{\partial\phi_j}=-\phi^{-1}(B)B^j a_t \quad , \quad \frac{\partial a_t}{\partial\Phi_j}=-\Phi^{-1}(B^s)B^{sj}a_t$$

如前一样，通过求矩阵 $I(\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{\Phi},\boldsymbol{\Theta})$ 的逆可以得到估计值的近似方差和协方差。

9.3.4 各种序列的最终预报函数

我们现在来考虑几个季节模型最终预报函数的特征。对于一个具有单一周期 s 的季节模型来说，在原点 t 提前期为 l 的最终预报函数是下面差分方程的解:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D\hat{z}_t(l)=0$$

表 9.8 给出了选择各种差分方程时的解，并给出了预报函数的特征所依赖初始值的数目.

在图 9.8 中对于 $s=4$ 说明了每个预报函数的特征. 把提前期 $l=rs+m$ 看作预报是提前 r 年 m 季度所做，这样是很方便的。在图中，适当数目的初值 (着手进行预报所要求的，用大黑点标出) 是任意设置的，预报函数的轨线一直持续到第四个周期末。当差分方程包含一个自回归参数时，它的值被设为等于 0.5.

在表 9.8 的解中出现的常数 $b_{0,m},b_1$ 等都应严格记作 $b_{0,m}^{(t)},b_1^{(t)}$ 等等，因为每一个常数都依赖于预报原点 t, 且随每一时刻原点的改变而被自适地修正。为了简化符号，上标 t 暂时省略。标号为

表 9.8 各种自回归算子的最终预报函数

自回归算子	最终预报函数 $\hat{z}(r,m)^*$	预报函数所依赖的初值个数
(1) $1-\Phi B^s$	$\mu-(b_{0,m}-\mu)\Phi^r$	s
(2) $1-B^s$	$b_{0,m}$	s
(3) $(1-B)(1-\Phi B^S)$	$b_0-(b_{0,m}-b_0)\Phi^r+b_1\left\{\frac{1-\Phi^r}{1-\Phi}\right\}$	$s+1$
(4) $(1-B)(1-B^S)$	$b_{0,m}+b_1r$	$s+1$
(5) $(1-\phi B)(1-B^S)$	$b_{0,m}+b_1\phi^{m-1}\{\frac{1-\phi^{sr}}{1-\phi^s}\}$	$s+1$
(6) $(1-B)(1-B^S)^2$	$b_{0,m}+b_{1,m}r+\frac{1}{2}b_2r(r-1)$	$2s+1$
(7) $(1-B)^2(1-B^S)^2$	$b_{0,m}+[b_1+(m-1)b_2]r$	$2s+$
	$+\frac{1}{2}b_2sr(r-1)$	$s+2$

* 系数 b 都是自适应的且依赖于预报原点 t .

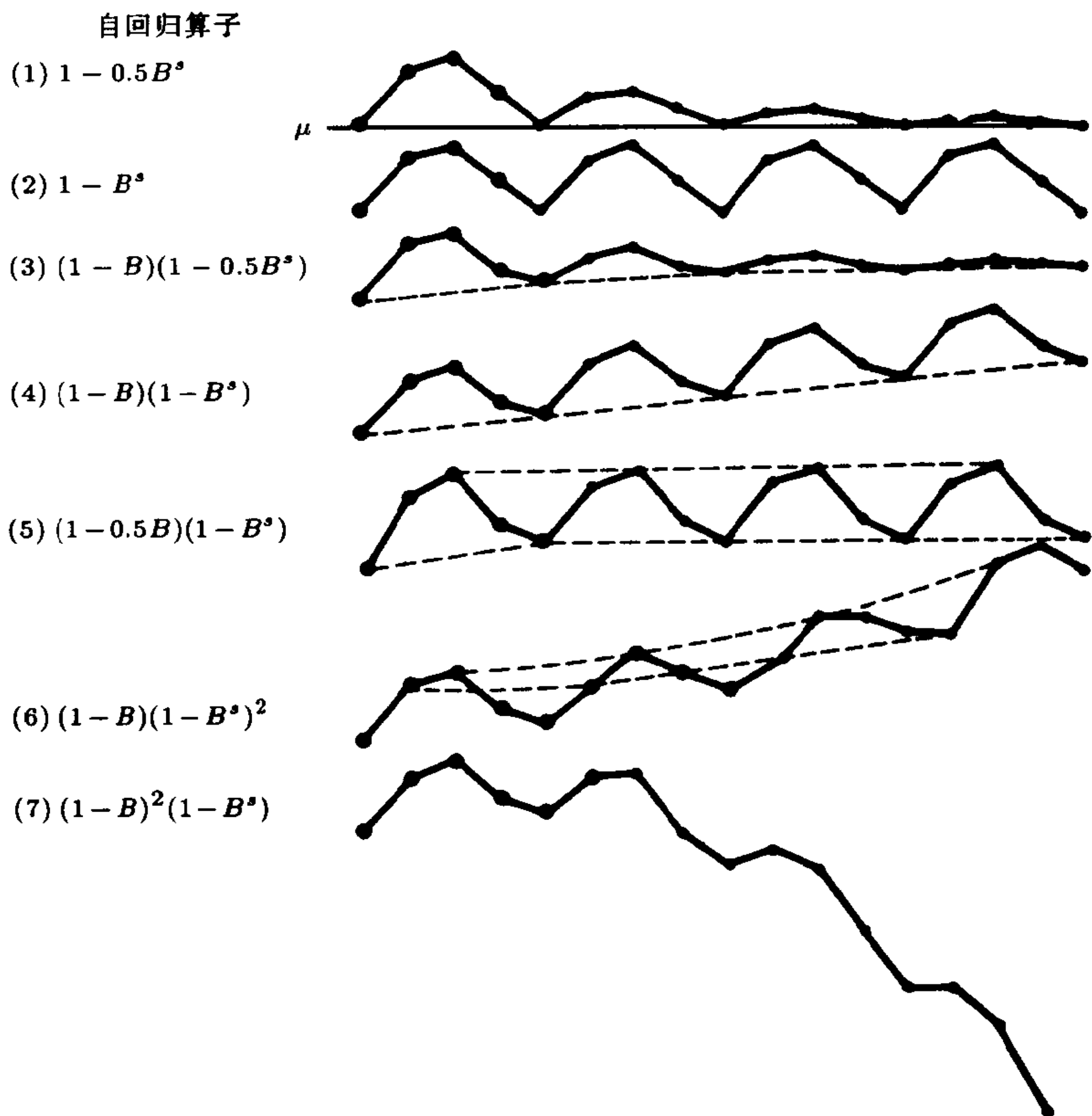


图 9.8 对于选择的各种一般季节自回归算子其季节预报函数的特征

(1) 的算子是平稳的，并包含一个固定的均值，它是自回归的季节模式，随着每个周期衰减，且越来越接近于均值。

具有季节分量的算子 (2) 是非平稳的。某个特定季度的预报值由一个零阶多项式将年与年之间联系起来。因此，季节分量的基本预报值在未来年份的预报中将得到完全地再现。

算子 (3) 关于基本间隔是非平稳的，但季节分量是平稳的。

图 9.8(3) 给出了预报值的一般水平，它渐近地接近于新水平

$$b_0 + \frac{b_1}{1-\Phi}$$

这里，迭加了平稳季节效应的可预报分量也同时指数衰减。

算子 (4) 是算子 (3) 当 Φ 接近于 1 时的极限情形。这个算子关于基本间隔和季节分量都是非平稳的。这个基本的初始预报模式象逐年增加的增量一样被再现。这就是拟合航空旅客数据的 $(0,1,1)\times(0,1,1)_{12}$ 乘积过程所给出的一类预报函数。

算子 (5) 关于季节模式是非平稳的，但关于基本间隔是平稳的。其形式指数地逼近下面的渐近形式

$$\hat{z}_t(\infty, m) = b_{0,m} + \frac{b_1\phi^{m-1}}{1-\phi^s}$$

算子 (6) 在基本间隔和季节分量上都是非平稳的. 在年度上呈现总体的二次趋势, 在季节模式中存在一种特定的修正. 每个季度不仅有自己的水平 $b_{0,m}$, 而且也有它们自己的水平变化率 $b_{1,m}$ 。因此, 如果这种预报函数是适合的, 我们便可能遇到这样的情形: 例如，随着提前期的增加，夏季对春季销售额之差可以预报为逐年增加，而同一时间秋季对夏季销售额之差却可能预报为下降。

算子 (7) 在基本间隔和季节分量上也都是非平稳的，仍然是存在年度上的二次趋势，而预报值从一个季度到下一季度的增长变化呈线性地改变。

9.3.5 变换的选择

用过去数据的加权平均来构成预报值，可以回溯到序列很久远的值，对于季节模型来说尤其如此。因此在选择变换形式时需要小心，通过此变换可以在该序列充分长的部分近似地应用一个简约的线性模型。简单的图形分析常常就可给出这样的变换。通过确定如何度量季节分量的幅度，且它与序列水平的度量大致是

独立的，由此可以得到一个适合的变换。为了说明怎样才能更精确地选择数据的基本变换，用 x 记未变换的航空旅客数据，并假设为使模型 (9.2.1) 适用，我们需要某种幂变换 ($z = x^{\lambda}$ 对 $\lambda \neq 0$, $z = \ln(x)$ 对 $\lambda = 0$)。因此，如 4.1.3 节所建议的，按照 Box 和 Cox[38] 的方法，对于各种 λ，对 $x^{(\lambda)} = (x^{\lambda} - 1)/\lambda \dot{x}^{\lambda - 1}$ 拟合模型，并选择使残差平方和 S_{λ} 最小的 λ 值来求得极大似然函数值。上面表达式中的 $\dot{x}$ 是序列的几何平均，很容易证明 $x^{(0)} = \dot{x}\ln(x)$。

对于航空旅客数据我们求出

λ	S_{λ}	λ	S_{λ}	λ	S_{λ}
−0.4	13,825.5	−0.1	11,627.2	0.2	11,784.3
−0.3	12,794.6	0	11,458.1	0.3	12,180.0
−0.2	12,046.0	0.1	11,554.3	0.4	12,633.2

这样，极大似然估计接近于 $\hat{\lambda} = 0$，对于这个例子来说，确认了对数变换的正确性。

9.4 结构分量模型和确定性季节分量

正如 9.1.1 节所提到的，描述季节时间序列的一种传统方法是把序列分解为趋势、季节及噪声分量，即 $z_t = T_t + S_t + N_t$，这里趋势 T_t 和季节 S_t 分别是采用多项式和正弦函数的时间的确定性函数。近来这种“结构分量”模型中含有的趋势和季节分量已不再是确定性的，而是遵从随机模型，这种模型已经在时间序列的建模，预报及季节调整等方面越来越流行 (例如 Harvey 和 Todd[109],Gersch 和 Kitagawa[95],Kitagawa 和 Gersch[128]，以及 Hillmer 和 Tiao[111])。例如对于月度数据，用这种模型处理趋势可假设服从模型 $(1 - B)T_t = (1 - \theta_T B)a_t$，或 $(1 - B)^2 T_t = (1 - \theta_{T1}B - \theta_{T2}B^2)a_t$，而假设季节服从 $(1 - B^{12})S_t = b_t$，这里 a_t 和 b_t 是独立白噪声过程。这种建模方法 —特别是对于季节调整问题 — 其 吸引力在于可以使用 5.5 节所讨论的基于模型状态空间公式的 Kalman 滤波和平滑方法。

然而应该注意，这样的结构模型具有等价的 ARIMA 模型表示。为说明起见，考虑 $z_t = T_t + S_t + N_t$，其中假设

$$(1-B)T_t = (1-\theta_T B)a_t\ , \qquad (1-B^{12})S_t = (1-\Theta_S B^{12})b_t$$

且 $N_t = c_t$ 是白噪声。于是我们有

$$\begin{aligned}(1-B)(1-B^{12})z_t =&(1-B^{12})(1-\theta_T B)a_t\\ &+(1-B)(1-\Theta_S B^{12})b_t + (1-B)(1-B^{12})c_t\end{aligned}$$

按照附录 A4.3 中的推广，上面表达式的右边可以表示为 MA 模型 $(1-\theta_1 B - \theta_{12}B^{12} - B_{13}B^{13})\epsilon_t$，这里 ϵ_t 是白噪声，这是因为右边仅在滞后为 0,1,11,12 和 13 时有非零的自协方差。在另外的结构之下 MA 算子可能有乘法形式，但是通常我们看到，如上述的结构模型一样，$z_t = T_t + S_t + N_t$ 具有等价的 ARIMA 表示如下

$$(1-B)(1-B^{12})z_t = (1-\theta_1 B - \theta_{12}B^{12} - \theta_{13}B^{13})\epsilon_t$$

9.4.1 确定的季节和趋势分量及公因子

在某些场合，特别是对于物理学中出现的序列，可以近似地有季节或趋势分量为确定性的。例如，假设季节分量事实上可表示为

$$S_t = \beta_0 + \sum_{j=1}^{6}\left[\beta_{1j}\cos\left(\frac{2\pi jt}{12}\right) + \beta_{2j}\sin\left(\frac{2\pi jt}{12}\right)\right]$$

这里系数 β 是确定性的 (固定的)。我们注意到，这可以视为前面例子的特殊情形，因为 S_t 满足 $(1-B^{12})S_t = 0$。现在，暂时忽略趋势项，在模型 $z_t = S_t + N_t$ 中假设 $(1-B^{12})S_t = 0$ 及 $N_t = (1-\theta_N B)c_t$，我们发现 z_t 服从下面的季节 ARIMA 模型

$$(1-B^{12})z_t = (1-\theta_N B)(1-B^{12})c_t$$

我们现在注意到，在该模型的广义 AR 算子和 MA 算子中都存在公因子 $1-B^{12}$，等价地，我们可以认为对于季节 MA 算子 $\Theta(B^{12})=(1-\Theta B^{12})$ 有 $\Theta=1$。这是由最初的模型形式中确定性（季节）分量 S_t 的存在所引起的，并且在事实上表示着它的存在。

通常，在一个时间序列 z_t 的结构中若存在确定性的季节或趋势分量，那么，在模型的广义 AR 算子和 MA 算子中都会显著地有公因子 $(1-B^{12})$ 或 $(1-B)$。我们可将这一结果更正式地如下文所叙述。假设 z_t 服从模型 $\varphi(B)z_t=\theta_0+\theta(B)a_t$，算子 $\phi(B)$ 和 $\theta(B)$ 包含公因子 $G(B)$，即 $\varphi(B)=G(B)\varphi_1(B)$ 及 $\vartheta(B)=G(B)\theta_1(B)$。因此，模型为

$$G(B)\varphi_1(B)z_t=\theta_0+G(B)\theta_1(B)a_t \tag{9.4.1}$$

令 $G(B)=1-g_1B-\cdots-g_rB^r$，并假设该多项式有根 $G_1^{-1},\cdots,G_r^{-1}$（假设是不等的）。于是公因子 $G(B)$ 可以从上面模型的两边去掉，但是要把形如 $\sum_{i=1}^{r}c_iG_i^r$ 的项加到模型上。于是模型 (9.4.1) 可用等价形式表示为

$$\varphi_1(B)z_t=c_{0t}+\sum_{i=1}^{t}c_iG_i^t+\theta_1(B)a_t \tag{9.4.2}$$

这里，c_i 是常数，项 c_{0t} 满足 $G(B)c_{0t}=\theta_0$。对于某些根 G_j^{-1} 为重根的情形，这个结果的修正是显而易见的。

因此可以看出，上面模型的一个等价表示为 $\varphi_1(B)z_t=x_t+\theta_1(B)a_t$，这里 x_t 是 t 的确定性函数，并满足 $G(B)x_t=\theta_0$。注意，$G(B)$ 的根对应于“平稳因子”，使得 $|G_i|<1$，故对分量 x_t 所做贡献是无关紧要的，因此这些项可以被忽略。这样，只有那些根对应于非平稳“差分”和其它“简化”算子的因子，例如 $(1-B)$ 和 $(1-B^s)$，其根为 $|G_i|=1$，只有这些因子才需要包含到模型等价形式的确定性分量中。当然，这些公因子会引出 x_t 中的确定性

函数，其形式为多项式，正弦和余弦函数及它们的乘积，这取决于公因子 $G(B)$ 的根。

例：对于一些简单的例子，模型 $(1-B)z_t=\theta_0+(1-B)\theta_1(B)a_t$ 具有等价形式 $z_t=c_1+\theta_0 t+\theta_1(B)a_t$，当然这是在消去公因子后出现的，而模型 $(1-\sqrt{3}B+B^2)z_t=\theta_0+(1-\sqrt{3}B+B^2)\theta_1(B)a_t$ 具有等价形式 $z_t=c_0+c_1\cos(2\pi t/12)+c_2\sin(2\pi t/12)+\theta_1(B)a_t$，这里 $(1-\sqrt{3}+1)c_0=\theta_0$。

在实际中，可能出现有关单位圆上根的公因子及相应约去因子的情形都只是近似的，然而，通过回归模型和 ARIMA 时间序列模型相结合的方法，直接去估计带有确定性分量 (隐含因子约去) 的试验性模型是很有用的。对于预报而言，存在确定性因子的另一后果是：在 (5.3.3) 中对 z_{t+l} 的一般预报函数 $\hat{z}_t(l)$ 中至少有一些系数 $b_j^{(t)}$ 不是自适应的，而是确定性 (固定) 的常数。前文中所讨论的结果，包括在 ARIMA 模型的广义 AR 算子和 MA 算子中具有单位圆上根的公因子，以及存在确定性的多项式和正弦分量，涉及二者之间的关系的讨论由 Abraham 和 Box[2],Harvey[106]，以及 Bell[27] 给出。

9.4.2 带有回归项和时间序列误差项的模型

前述讨论引发了一类时间序列模型的讨论 — 它包含诸如确定性的正弦余弦函数的回归项，用以描述季节特性，并包含有可能自相关的时间序列误差 (或"噪声") 项。我们考虑误差项 N_t 可以假设为是平稳 ARMA 过程的情形，否则，我们就可能要考虑对原序列进行差分。所以，我们假设序列 w_t 服从回归模型

$$w_t=\beta_1 x_{t1}+\beta_1 x_{t2}+\cdots+\beta_k x_{tk}+N_t \qquad (t=1,\cdots,n) \quad (9.4.3)$$

这里，$x_{t1},\cdots,x_{tk}$ 是 k 个解释变量或预测变量，误差 N_t 假设是由一个 $\text{ARMA}(p,q):\phi(B)N_t=\theta(B)a_t$ 生成。回忆一下传统线性回归模型的结果，在附录 A7.2 中对此做了简要回顾。用那里的符号，模型 (9.4.3) 可以写为 $\boldsymbol{w}=\boldsymbol{X\beta}+\boldsymbol{N}$，及 $\boldsymbol{V}=\text{cov}[\boldsymbol{N}]$。在传统模

型中我们有 $\boldsymbol{V}=\sigma^2\boldsymbol{I}$, 且 $\boldsymbol{\beta}$ 的最小二乘估计为 $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{w}$, 并具有通常的性质，如 $\text{cov}[\hat{\boldsymbol{\beta}}]=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

但是在自相关误差的情形，这种性质不再成立，我们发现最小二乘估计的协方差阵是由 $\text{cov}[\hat{\boldsymbol{\beta}}]=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$. 因此一般的样本性质和统计推断方法，例如关于估计值 $\hat{\beta}_i$ 的通常标准差公式， t- 统计量及置信区间等就不再有效了，因为它们都是建立在关于 (9.4.3) 中误差 N_t 性质的不正确假设之上，由此会导致最小二乘估计值 $\hat{\boldsymbol{\beta}}$ 协方差阵的不正确形式。

例：作为说明这种情况的简单例子，考虑模型 $w_t=\beta_1 x_t+N_t$, 这里 N_t 服从 MA(1) 模型 $N_t=(1-\theta B)a_t$ 。于是我们知道， N_t 仅有 $\gamma_0=(1+\theta^2)\sigma_a^2$ 和 $\gamma_1=-\theta\sigma_a^2$ 是非零的自协方差。 β_1 的最小二乘估计是 $\hat{\beta}_1=\sum\limits_{t=1}^{n}x_t w_t/\sum\limits_{t=1}^{n}x_t^2$, 它的实际方差为

$$\text{var}[\hat{\beta}_1]=\frac{\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}}{(\sum\limits_{t=1}^{n}x_t^2)^2}=\frac{\gamma_0\sum\limits_{t=1}^{n}x_t^2+2\gamma_1\sum\limits_{t=1}^{n-1}x_t x_{t+1}}{(\sum\limits_{t=1}^{n}x_t^2)^2}$$

该表达式可简化为

$$\text{var}[\hat{\beta}_1]=\frac{\gamma_0}{\sum\limits_{t=1}^{n}x_t^2}[1+2\rho_1 r_1(x)]$$

这里 $r_1(x)=\sum\limits_{t=1}^{n-1}x_t x_{t+1}/\sum\limits_{t=1}^{n}x_t^2$ 可解释为 x_t 滞后为 1 的样本自相关函数，不过未对均值修正。与此对照，如果模型中误差 N_t 解释为不相关的，我们会取 $\rho_1=0$, 因而不正确地导致认为 $\hat{\beta}_1$ 的方差是 $\gamma_0/\sum\limits_{t=1}^{n}x_t^2$ 。而这种形式方差的应用会引出有关 β_1 的错误推断。在这个简单的例子中我们还看到，自相关 ρ_1 对最小二乘估计方差的影响是通过因子 $\rho_1 r_1(x)$ 而依赖于解释变量 x_t 的自相关特征。因此，当序列 x_t 和噪声 N_t 有相同符号的 1 阶滞后自相关时，最

小二乘估计的方差将会比具有不相关误差时扩大，而当具有相反符号的自相关时，结果相反。

当对时间序列数据拟合回归模型时，总是需要考虑误差项存在自相关的可能性. 通常，对于误差 N_t 识别一个恰当模型的可行方法是首先得出最小二乘估计 $\hat{\boldsymbol{\beta}}$, 然后再得到相应回归模型的残差

$$\hat{N}_t = w_t - \hat{\beta}_1 x_{t1} - \hat{\beta}_2 x_{t2} - \cdots - \hat{\beta}_k x_{tk} \tag{9.4.4}$$

这个残差序列可以用通常的时间序列建模方法来考察，例如考察它的样本自相关和偏相关函数，以对 N_t 识别合适的 ARMA 模型。这种预先的识别技巧足以对误差项 N_t 建立一个试验性的模型。尤其是当自变量 x_{ti} 为确定性函数形式, 比如正弦、余弦函数, 多项式等，因为在这些情形下我们知道 (例如见 Anderson[8,10.2 节]), 最小二乘估计相对于最优线性估计是渐近有效的估计量。另外，我们还知道，由预先最小二乘拟合残差计算的样本自相关和偏关与从实际噪声序列 N_t 所得的相应值是渐近相等的 (例如见 Anderson[8 中 10.3 节] 和 Fuller[93 的 9.3 节]) 。一般， $\boldsymbol{\beta}$ 的最优线性估计，通称广义最小二乘估计由 $\hat{\boldsymbol{\beta}}_G = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{w}$ 给出，其协方差阵为 $\text{cov}[\hat{\boldsymbol{\beta}}_G] = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ 。当协方差阵 $\boldsymbol{V}$ 已知，且误差为正态的假定下，该估计量也对应于极大似然估计。

在实际中，根据对噪声项 N_t 确定的具体 ARMA 模型，可以确定 $\boldsymbol{N}$ 的协方差阵 $\boldsymbol{V}$ 的具体形式, 并可找到一个下三角阵 $\boldsymbol{P}'$, 使得 $\boldsymbol{P}'\boldsymbol{V}\boldsymbol{P} = \sigma_a^2\boldsymbol{I}$, 即 $\boldsymbol{V}^{-1} = \boldsymbol{P}\boldsymbol{P}'/\sigma_a^2$ 。因此，如 A7.2.5 节所述，事实上, 可以由变换后的变量 $\boldsymbol{w}^* = \boldsymbol{P}'\boldsymbol{w}$ 及 $\boldsymbol{X}^* = \boldsymbol{P}'\boldsymbol{X}$ 得到广义最小二乘估计 $\hat{\boldsymbol{\beta}}_G = (\boldsymbol{X}^{*'}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*'}\boldsymbol{w}^*$, 且有 $\text{cov}[\hat{\boldsymbol{\beta}}_G] = \sigma_a^2(\boldsymbol{X}^{*'}\boldsymbol{X}^*)^{-1}$ 。因为 N_t 的时间序列模型参数 ϕ_i 和 θ_i 是未知的，所以必须在计算 $\hat{\boldsymbol{\beta}}_G$ 和极大似然估计 (或精确最小二乘法) 之间迭代，前者是利用时间序列模型参数的当前估计值以构造变换矩阵 $\hat{\boldsymbol{P}}'$, 后者则如第 7 章所述，由 $\boldsymbol{\beta}$ 的当前估计值构造出估计的噪声序列 $\hat{N}_t$, 在此基础上得出时间序列模型参数的估计值。用于对指定 ARMA 模型

确定精确平方和函数的计算过程，也将基本上确定计算变换矩阵 $\boldsymbol{P}'$ 的性质，该矩阵是计算广义最小二乘估计值所需要的。

例： 我们用简单的例子加以说明，对噪声 N_t 取 AR(1) 模型： $(1-\phi B)N_t = a_t$ 。于是 $n\times n$ 阵 $\boldsymbol{P}'$ 的 (1,1) 元素等于 $(1-\phi^2)^{1/2}$ ，其余对角线元素都等于 1，主对角线下的对角线元素等于 $-\phi$，其余所有元素都是零。因此在该变换之后变量成为 $w_1^* = (1-\phi^2)^{1/2}w_1$，及 $w_t^* = w_t - \phi w_{t-1}, t=2,3,\cdots,n$，对于变换后的解释变量 x_{ti}^* 也可类似得到。实际上，有了 AR(1) 误差，通过对方程运用算子 $(1-\phi B)$，模型 (9.4.3) 变换后得到

$$\begin{aligned} w_t - \phi w_{t-1} =& \beta_1(x_{t1} - \phi x_{t-1,1}) + \beta_2(x_{t2} - \phi x_{t-1,2}) \\ &+ \cdots + \beta_k(x_{tk} - \phi x_{t-1,k}) + a_t \end{aligned}$$

或 $w_t^* = \beta_1 x_{t-1}^* + \beta_2 x_{t2}^* + \cdots + \beta_k x_{tk}^* + a_t$，这是带有不相关误差 a_t 的通常回归模型形式，因此普通最小二乘回归方法适用于变换后的模型。对于高阶 AR 模型来说，变换方法的推广是直截了当的，这不同于对于 p 个初始观测值的特殊处理，于是变换后变量为 $x_{ti}^* = \phi(B)x_{ti} = x_{ti} - \phi_1 x_{t-1,i} - \cdots - \phi_p x_{t-p,i}, i=1,\cdots,k$ 。对混合 ARMA 模型的情形，变换的精确形式将更复杂 (一种近似形式是 $w_t^* \simeq \theta^{-1}(B)\phi(B)w_t$，等等)，但是正如所指出的，确定这些变换所用的方法是和用来构造 ARMA 模型精确平方和函数的方法一样的。

当解释变量的未来值 $x_{t+l,i}$ 假设是已知的，那么，对于带有时间序列误差的回归模型的预报就是显而易见的了，这就象是对确定性函数 (例如正弦或余弦函数) 一样。于是，若预报原点 t，提前期为 l 时，

$$w_{t+l} = \beta_1 x_{t+l,1} + \cdots + \beta_k x_{t+l,k} + N_{t+l}$$

基于直到时刻 t 过去值的预报为

$$\hat{w}_t(l) = \beta_1 x_{t+l,1} + \beta_2 x_{t+l,2} + \cdots + \beta_k x_{t+l,k} + \hat{N}_t(l) \qquad (9.4.5)$$

这里 $\hat{N}_t(l)$ 是由 ARMA(p,q) 模型 $\phi(B)N_t=\theta(B)a_t$ 给出、以噪声序列 N_t 的过去值为基础对 N_{t+l} 提前 l 期的通常预报值. 预报误差为

$$e_t(l)=w_{t-l}-\hat{w}_t(l)=N_{t+l}-\hat{N}_t(l)=\sum_{i=0}^{l-1}\psi_i a_{t+l-i} \qquad (9.4.6)$$

并有 $V(l)=\text{var}[e_t(l)]=\sigma_a^2\sum_{i=0}^{l-1}\psi_i^2$, 这是噪声序列 N_t 的 ARMA 模型预报误差及方差, 其中 ψ_i 是噪声模型中 $\psi(B)=\phi(B)^{-1}\theta(B)$ 的系数。

例：作为例子, 在模型 $w_t=\beta_0+\beta_1\cos(2\pi t/12)+\beta_2\sin(2\pi t/12)+N_t$ 中, 这里 $(1-\phi B)N_t=a_t$, 预报值是

$$\hat{w}_t(l)=\beta_0+\beta_1\cos\left[\frac{2\pi(t+l)}{12}\right]+\beta_2\sin\left[\frac{2\pi(t+l)}{12}\right]+\hat{N}_t(l)$$

其中 $\hat{N}_t(l)=\phi^l N_t$. 注意, 这些预报值在函数形式方面类似于用一个 ARMA(1,3) 模型 (常数项为 0) 对序列 $(1-B)(1-\sqrt{3}B+B^2)w_t$ 得到的预报值, 正如 9.4.1 节末尾所指出的, 只是对于回归模型情形, 预报函数的 β 系数是确定性的, 而不是自适应的.

在实际中, 与时间序列模型的参数估计值一样, $\boldsymbol{\beta}$ 的估计值可用来求得估计的噪声序列 $\hat{N}_t$, 以此可作出对未来值的预测. 有关这些参数估计误差对相应预报误差方差的影响, 由 Baillie[20] 对带有自回归误差的回归模型情形作了研究, Yamamoto[210] 关于参数估计误差对纯 AR 模型预报的影响做了更一般的研究。

对于带有时间序列误差的回归模型方法更详细的讨论由 Harvey 和 Phillips[107] 以及 Wincek 和 Reinsel[205] 给出, 后者还涉及缺失数据的可能性。在 7.5 节中所讨论的用 Kalman 滤波计算的状态空间方法可应用于带有时间序列误差的回归模型, 并且关于前面所讨论的变换矩阵 $\boldsymbol{P}'$, 在这里相应于一种特定的选择。作为特殊的应用, Bell 和 Hillmer[28] 将带有时间序列误差的回归模

型用于季节时间序列中日历影响的建模， Reinsel 和 Tiao[167] 用带有时间序列误差的回归模型对空气中臭氧的趋向估计建模.

我们用引自 [167] 的一个例子作为说明，考虑序列 z_t, 它是澳大利亚 Aspendale 站观测的大气中臭氧的月平均值，时间为 1958-1984. 该序列有很强的季节性，因此用 ARIMA 建模，考虑季节差分 $w_t = (1-B^{12})z_t$, 以 w_t 样本的 ACF 和 PACF 特征为基础，给出并估计下面的模型

$$(1-0.48B-0.22B^2)(1-B^{12})z_t = (1-0.99B^{12})a_t$$

该模型经断定是恰当的. 因此，我们看到模型中含有十分接近的季节差分公因子 $(1-B^{12})$, 从而，它等价于包含确定性季节分量的模型

$$z_t = \beta_0 + \sum_{j=1}^{6}\left[\beta_{1j}\cos\left(\frac{2\pi jt}{12}\right) + \beta_{2j}\sin\left(\frac{2\pi jt}{12}\right)\right] + N_t$$

这里， N_t 服从 AR(2) 模型： $(1-0.48B-0.22B^2)N_t = a_t$. 事实上，这一模型形式在 [167] 中是用回归模型 - 时间序列方法估计出的，这类似于我们前面所讨论的方法.

A9.1 某些季节模型的自协方差

见表 A9.1 .

表 A9.1 某些季节模型的自协方差

模型	w_t 的自协方差 $1/\sigma_a^2$	特征
(1) $w_t=(1-\theta B)(1-\Theta B^S)a_t$	$\gamma_0=(1+\theta^2)(1+\Theta^2)$	(a) $\gamma_{S-1}=\gamma_{S+1}$
$w_t=a_t-\theta a_{t-1}-\Theta a_{t-S}$	$\gamma_1=-\theta(1+\Theta^2)$	(b) $\rho_{S-1}=\rho_{S+1}$
$+\theta\Theta a_{t-S-1}$	$\gamma_S=-\Theta(1+\theta^2)$	$=\rho_1\rho_S$
$S>3$	$\gamma_{S-1}=\theta\Theta$，$\gamma_{S+1}=\gamma_{S-1}$	
	所有其它自协方差皆为零	
(2) $(1-\Phi B^S)w_t=$	$\gamma_0=(1+\theta^2)\left[1+(\Theta-\Phi)^2/(1-\Phi^2)\right]$	(a) $\gamma_{S-1}=\gamma_{S+1}$
$(1-\theta B)(1-\Theta B^S)a_t$	$\gamma_1=-\theta\left[1+(\Theta-\Phi)^2/(1-\Phi^2)\right]$	(b) $\gamma_j=\Phi\gamma_{j-S}$
$w_t-\Phi w_{t-S}=a_t-\theta a_{t-1}$	$\gamma_{S-1}=\theta[\Theta-\Phi-\Phi(\Theta-\Phi)^2/(1-\Phi^2)]$	$j>s+2$
$-\Theta a_{t-S}+\theta\Theta a_{t-S-1}$	$\gamma_S=-(1+\theta^2)$	
	$\times[\Theta-\Phi-\Phi(\Theta-\Phi)^2/(1-\Phi^2)]$	
$S>3$	$\gamma_{S+1}=\gamma_{S-1}$	
	$\gamma_j=\Phi\gamma_{j-S}$，$j>S+2$	
	对于 $S>4,\gamma_2,\gamma_3,\cdots,\gamma_{S-2}$ 皆为零	
(3) $w_t=(1-\theta_1 B-\theta_2 B^2)$	$\gamma_0=(1+\theta_1^2+\theta_2^2)(1+\Theta_1^2+\Theta_2^2)$	(a) $\gamma_{S-2}=\gamma_{S+2}$
$\times(1-\Theta_1 B^S-\Theta_2 B^{2S})a_t$	$\gamma_1=-\theta_1(1-\theta_2)(1+\Theta_1^2+\Theta_2^2)$	(b) $\gamma_{S-1}=\gamma_{S+1}$
$w_t=a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2}$	$\gamma_2=-\theta_2(1+\Theta_1^2+\Theta_2^2)$	(c) $\gamma_{2S-2}=\gamma_{2S+2}$
$-\Theta_1 a_{t-S}+\theta_1\Theta_1 a_{t-S-1}$	$\gamma_{S-2}=\theta_2\Theta_1(1-\Theta_2)$，$\gamma_{S+2}=\gamma_{S-2}$	(d) $\gamma_{2S-1}=\gamma_{2S+1}$
$-\Theta_2 a_{t-2S}+\theta_2\Theta_1 a_{t-S-2}$	$\gamma_{S-1}=\theta_1\Theta_1(1-\theta_2)(1-\Theta_2)$	
$+\theta_1\Theta_2 a_{t-2S-1}$	$\gamma_S=-\Theta_1(1+\theta_1^2+\theta_2^2)(1-\Theta_2)$	
$+\theta_2\Theta_2 a_{t-2S-2}$	$\gamma_{S+1}=\gamma_{S-1}$	
$S>5$	$\gamma_{2S-2}=\theta_2\Theta_2$，$\gamma_{2S+2}=\gamma_{2S-2}$	

表 A9.1　某些季节模型的自协方差 (续 1)

模　型	w_t 的自协方差 $1/\sigma_a^2$	特　征
	$\gamma_{2S-1} = \theta_1\Theta_2(1-\theta_2)$ $\gamma_{2S} = -\Theta_2(1+\theta_1^2+\theta_2^2)$ $\gamma_{2S+1} = \gamma_{2S-1}$ 所有其它自协方差皆为零	
(3a) 模形 (3) 的特例 $w_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta B^S)+a_t$ $w_t = a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2}-\Theta a_{t-S}$ $+\theta_1\Theta a_{t-S-1}+\theta_2\Theta a_{t-S-2}$ $S > 5$	$\gamma_0 = (1+\theta_1^2+\theta_2^2)(1+\Theta^2)$ $\gamma_1 = -\theta_1(1-\theta_2)(1+\Theta^2)$ $\gamma_2 = -\theta_2(1+\Theta^2)$ $\gamma_{S-2} = \theta_2\Theta$, $\gamma_{S+2} = \gamma_{S-2}$ $\gamma_{S-1} = \theta_1\Theta(1-\theta_2)$, $\gamma_S = -\Theta(1+\theta_1^2+\theta_2^2)$ $\gamma_{S+1} = \gamma_{S-1}$ 所有其它自协方差皆为零	(a) $\gamma_{S-2} = \gamma_{S+2}$ (b) $\gamma_{S-1} = \gamma_{S+1}$
(3b) 模形 (3) 的特例 $w_t = (1-\theta B)(1-\Theta_1 B^S-\Theta_2 B^{2S})a_t$ $w_t = a_t-\theta_1 a_{t-1}-\Theta a_{t-S}+\theta\Theta_1 a_{t-S-1}$ $-\Theta_2 a_{t-2S}+\theta\Theta_2 a_{t-2S-1}$ $S > 3$	$\gamma_0 = (1+\theta_2)(1+\Theta_1^2+\Theta^2)$ $\gamma_1 = -\theta(1+\Theta_1^2+\Theta_2)$ $\gamma_{S-1} = \theta\Theta_1(1-\Theta_2)$ $\gamma_S = -\Theta_2(1-\theta^2)(1-\Theta_2)$ $\gamma_{S+1} = \gamma_{S-1}$ $\gamma_{2S-1} = \theta\Theta_2$, $\gamma_{2S+1} = \gamma_{2S-1}$ $\gamma_{2S} = -\Theta_2(1+\theta^2)$ 所有其它自协方差皆为零	(a) $\gamma_{S-1} = \gamma_{S+1}$ (b) $\gamma_{2S-1} = \gamma_{2S+1}$

表 A9.1 某些季节模型的自协方差 (续 2)

模 型	w_t 的自协方差 $1/\sigma_a^2$	特 征 •
(4) $w_t=(1-\theta_1 B-\theta_S B^S -\theta_{S+1}B^{S+1})a_t$ $w_t=a_t-\theta_1 a_{t-1}-\theta_S a_{t-S}-\theta_{S+1}a_{t-S-1}$ $S>3$	$\gamma_0=1+\theta_1^2+\theta_S^2+\theta_{S+1}^2$ $\gamma_1=-\theta_1+\theta_S\theta_{S+1}$ $\gamma_{S-1}=\theta_1\theta_S$ $\gamma_S=\theta_1\theta_{S+1}-\theta_S$ $\gamma_{S+1}=-\theta_{S+1}$ 所有其它自协方差皆为零	(a) 一般有: $\gamma_{S-1}\neq\gamma_{s+1}$ $\gamma_1\gamma_S\neq\gamma_{S+1}$
(4a) 模型 (4) 的特例 $w_t=(1-\theta_1 B-\theta_S B^S)a_t$ $w_t=a_t-\theta_1 a_{t-1}-\theta_S a_{t-S}$ $S>3$	$\gamma_0=1+\theta_1^2+\theta_S^2$ $\gamma_1=-\theta_1$ $\gamma_{S-1}=\theta_1\theta_S$, $\gamma_S=-\theta_S$ 所有其它自协方差皆为零	(a) 与模型 (4) 不同 $\gamma_{S+1}=0$
(5) $(1-\Phi B^S)w_t=(1-\theta_1 B-\theta_S B^S -\theta_{S+1}B^{S+1})a_t$ $w_t-\Phi w_{t-S}=a_t-\theta a_{t-1}-\theta_S a_{t-S}-\theta_{S+1}a_{t-S-1}$	$\gamma_0=1+\theta_1^2+\frac{(\theta_1-\Phi)^2}{1-\Phi^2)}+\frac{(\theta_{S+1}+\theta_1\Phi)^2}{1-\Phi^2}$ $\gamma_1=-\theta_1+\frac{(\theta_1-\Phi)(\theta_{S+1}+\theta_1\Phi)}{(1-\Phi^2)}$ $\gamma_{S-1}=(\theta_S-\Phi)[\theta_1+\Phi\frac{\theta_{S+1}-\Phi\theta_1}{1-\Phi^2}]$ $\gamma_S=-(\theta_S-\Phi)[1-\Phi\frac{\theta_S-\Phi}{1-\Phi^2}]$	(a) $\gamma_{S-1}\neq\gamma_{S+1}$ (b) $\gamma_j=\Phi\gamma_{j-1}$ $j>S+2$

表 A9.1 某些季节模型的自协方差 (续 3)

模 型	w_t 的自协方差 $1/\sigma_a^2$	特 征
$S > 3$	$+(\theta_{S+1}+\theta_1\Phi)[\theta_1+\Phi\frac{\theta_{S+1}+\theta_1\Phi}{1-\Phi^2}]$	
	$\gamma_{S+1}=-(\theta_{S+1}+\theta_1\Phi)[1-\Phi\frac{\theta_S-\Phi}{1-\Phi^2}]$	
	$\gamma_j=\Phi\gamma_{j-S}$, $j > S+2$	
	对于 $S > 4$, $\gamma_2,\cdots,\gamma_{S-2}$ 皆为零	
(5a) 模形 (5) 的特例	$\gamma_0=1+\frac{\theta_1^2+(\theta_S-\Phi)^2}{1-\phi^2}$	(a) 与模型 (5) 不同 $\gamma_{S+1}=\Phi\gamma_1$
$(1-\Phi B^S)w_t$ $=(1-\theta_1 B-\theta_S B^S)a_t$	$\gamma_1=-\theta_1[1-\Phi\frac{\theta_S-\Phi}{1-\Phi^2}]$	
$w_t-\Phi w_{t-S}$ $=a_t-\theta_1 a_{t-1}+\theta_S a_{t-S}$	$\gamma_{S-1}=\frac{\theta_1(\theta_S-\Phi)}{1-\Phi^2}$	
$S > 3$	$\gamma_S=\frac{\Phi\theta_1^2-(\theta_S-\Phi)(1-\Phi\theta_S)}{1-\Phi^2}$	
	$\gamma_j=\Phi\gamma_{j-s}$ $j > S+1$	
	对于 $S > 4$, $\gamma_2,\cdots,\gamma_{S-2}$ 皆为零	

第三部分

传递函数模型的建立

假设 X 用来度量系统输入的水平。例如，X 可以是馈入化学过程中某种成分的浓度。假定 X 的水平影响系统输出 Y 的水平，比如，Y 可以是化工过程的产出。情况往往是这样：由于系统的惯性，X 从一种水平变化到另一水平时，并不是立刻影响到输出，而是引起 Y 的滞后的响应，最后，在一个新的水平上达到均衡。我们把这种变化称为动态系统的响应。描述这种动态响应的模型叫做传递函数模型。我们将假设输入和输出的观测值都是在等间隔的时间上得到的，那么，相应的传递函数就称为离散的传递函数模型。

这类模型不仅能够描述工业过程的特性，而且也能用于经济和商业系统。传递函数模型的建立是很重要的。因为只有了解系统的动态特征，对于该系统进行明智的引导、操纵和控制才具有可能性。

即使在精心控制之下，除 X 以外的其它影响也会对 Y 有作用。我们把这些影响对 Y 的综合作用称为干扰或噪声。一个能够联系实际数据的模型不仅要考虑连接 X 和 Y 的动态关系，而且也要考虑影响系统的噪声。这种联合模型是通过把确定性的传递函数模型与随机噪声模型组合而得到的。

在下面的第 10 章中我们引入一族线性传递函数模型，它们可以表示实际中经常遇到的许多动态关系。在第 11 章中我们要说明应该怎样考虑噪声干扰，才可以使噪声与数据相联系。这种模型和数据的联系是通过识别、估计和诊断检验的过程来实现的，这和前面已经讲述的内容非常类似. 在第 12 章中，我们将描述：如何在传递函数模型中把简单脉冲和阶跃示性变量用作输入，以此来表示和评价非正常干预事件对时间序列 Y 特性的影响。

10
传递函数模型

本章中我们介绍一族离散的线性传递函数模型。它们可以用来表示经常遇到的动态情形，且模型中参数的使用是简约的。

10.1 线性传递函数模型

我们假定，如图 10.1 中所示，在相等的时间间隔上从某动态系统的输入 X 和输出 Y 可以得到一对观测值 (X_t, Y_t)。在一些场合，X 和 Y 实质上都是连续的，但仅在离散时刻上进行观测。我们不仅考虑描述将一个离散序列转换为另一个的模型 — 这是从数据必然得知的，而且还要考虑从离散模型可以得出的关于相应连续模型的情况，这是十分有意义的。在其它一些例子中，离散序列本身存在并不以连续过程为背景。当把连续系统和离散系

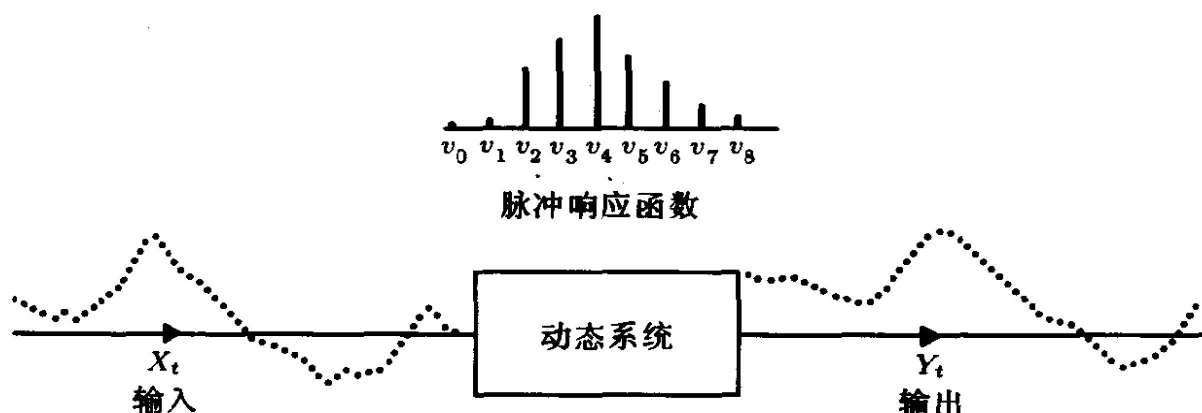


图 10.1　一个动态系统的输入和输出

统相联系时，我们就利用基本采应间隔作为时间单位。这就是说，时间的长度是由它所占采样间隔的数目来度量的，也即，一个离散观测值 X_t 被认为恰好是在“时刻 t”发生。

当我们考虑连续变量的值时，如 Y 在时刻 t 的值，我们用 $Y(t)$ 来记。如果 t 恰好是一个离散变量 Y 被观测的时刻，其值便记作 Y_t 。当我们要强调一个离散输出 Y 不仅依赖于时间，而且还依赖于输入 X 的水平时，我们写为 $Y_t(X)$ 。

10.1.1 离散传递函数

在适当的输入和输出之下，读者可以想象图 10.1 的动态系统是表示一个工业过程，一个国家的经济，或者是一个具体的公司或政府部门的状态。

当输入保持某固定值时，我们常常称所得到的输出为稳态水平。由此，我们把 $Y_\infty(X)$ 理解为：当输入保持固定水平 X 时，最终达到均衡状态的稳定系统的离散输出。通常在我们感兴趣的范围内， $Y_\infty(X)$ 和 X 之间的关系是近似于线性的。因此，如果我们用 Y 和 X 记从某适当原点到直线距离的偏差，那么，我们便可将稳态关系写为

$$Y_\infty = gX \tag{10.1.1}$$

这里 g 称为稳态增益，该式可以理解为 Y_∞ 是 X 的函数。

现在，假设输入水平是变化的，那么， X_t 和 Y_t 都表示时刻 t 自均衡状态的偏差。那么经常发生的情况是：在适当的近似下，系统的惯性可以用一个线性滤波器来表示：

$$\begin{aligned} Y_t &= v_0X_t + v_1X_{t-1} + v_2X_{t-2} + \cdots \\ &= (v_0 + v_1B + v_2B^2 + \cdots)X_t \\ &= v(B)x_t \end{aligned} \tag{10.1.2}$$

在上式中，某时刻 t 的输出偏差表示成在时刻 $t, t-1, \cdots$ 输入偏差的线性组合。算子 $v(B)$ 称为滤波器的传递函数。

脉冲响应函数 (10.1.2) 中的权数 $v_0, v_1, v_2, \cdots$ 称为系统的脉冲响应函数。这是因为 v_j 可以看作是：在时刻 $j \geqslant 0$ 对于时刻 0 的单位脉冲的输出或响应，即输入 X_t 在 $t=0$ 时 $X_t=1$; 其它时刻 $X_t=0$。脉冲响应函数在图 10.2 中以棒形图的形式给出。当没有瞬时响应时，一个或多个初始的 v 值，如 $v_0, v_1, \cdots, v_{b-1}$ 将等于零。

根据 (10.1.2), 输出偏差可以看作是一系列重叠脉冲响应函数乘以偏差 X_t 的线性组合。图 10.2 对此作了说明，图中给出了一个假设的脉冲响应函数，以及由它引起的从输入到输出的传递。在所示情形中，输入和输出在一开始是均衡的，在时刻 $t=1, t=2$ 和 $t=3$ 输入发生的偏差引起了输出的脉冲响应式的偏差，它们加在一起形成了整体的输出响应。

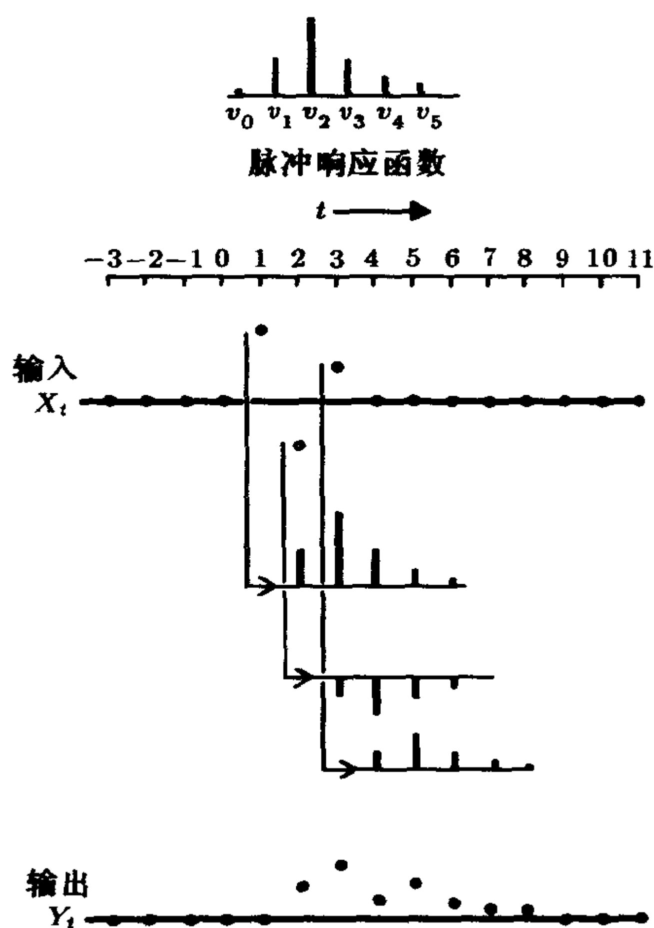


图 10.2 从输入到输出的线性传递

增量变化之间的关系　用

$$y_t = Y_t - Y_{t-1} = \nabla Y_t$$

和

$$x_t = X_t - X_{t-1} = \nabla X_t$$

来记 Y 和 X 的增量变化。我们常常希望把这些变化联系起来。对 (10.1.2) 差分，我们得到

$$y_t = v(B)x_t$$

由此我们看到，增量变化 y_t 和 x_t 满足与 Y_t 和 X_t 同样的传递函数模型。

稳定性　如果无穷级数 $v_0 + v_1B + v^2B^2 + \cdots$ 对于 $|B| < 1$ 收敛，或等价地，v_j 为绝对可和，即 $\sum_{j=0}^{\infty}|v_j| < \infty$，那么系统就被称为是稳定的。在这里我们只关心稳定系统，因此把这个条件加在我们所研究的模型之上。稳定性条件意味着输入的一个有限增量变化将导致输出的有限增量变化。

现在假定 X 永远保持在 $+1$ 这个值，那么根据 (10.1.1),Y 将调整并维持在值 g 。在 (10.1.2) 中令 $Y_t = g, 1 = X_t = X_{t-1} = X_{t-2} \cdots$，我们得到

$$\sum_{j=0}^{\infty} v_j = g \tag{10.1.3}$$

所以一个稳定系统的脉冲响应权数之和收敛且等于系统的稳态增益。

简约性　为使系统参数化，用 (10.1.2) 中 v 的表示通常是不能令人满意的。如此冗费地使用参数将会在估计阶段导致传递函数的不准确和不稳定估计。因为在许多实际场合，v 是函数值相关的，正象我们现在所见到的。

10.1.2 用微分方程表示的连续动态模型

一阶动态系统　考虑图 10.3。假设在时刻 t,$X(t)$ 是罐 A 中水的体积，$Y(t)$ 是罐 B 中水的体积，A 和 B 通过一条管道相连接。罐 C 是用虚线表示的，一开始我们先不管它。假设水可以通过管道 P 流入或流出 A, 这样，可以利用机械手段强制 A 中的水平面也即体积达到预期的状态，而不必管在 B 中发生了什么。

现在，如果第一个罐中的体积 X 控制在某个固定水平，水将会从一个罐流到另一个罐，直到两者水平面相同。如果我们现在重新调整体积 X 到另一个值，两个水罐之间的流动又会发生直到实现均衡。B 中的均衡体积是 A 中固定体积的函数，由此得出稳态关系

$$Y_{1\infty} = g_1 X \tag{10.1.4}$$

在这个例子中，稳态增益 g_1 在物理上表示两个罐的横截面积之比。如果在某一时刻 t 两者的水平尚未达到均衡，可以注意到：两个罐的水平面之差与 $X(t) - Y_1(t)$ 成比例。

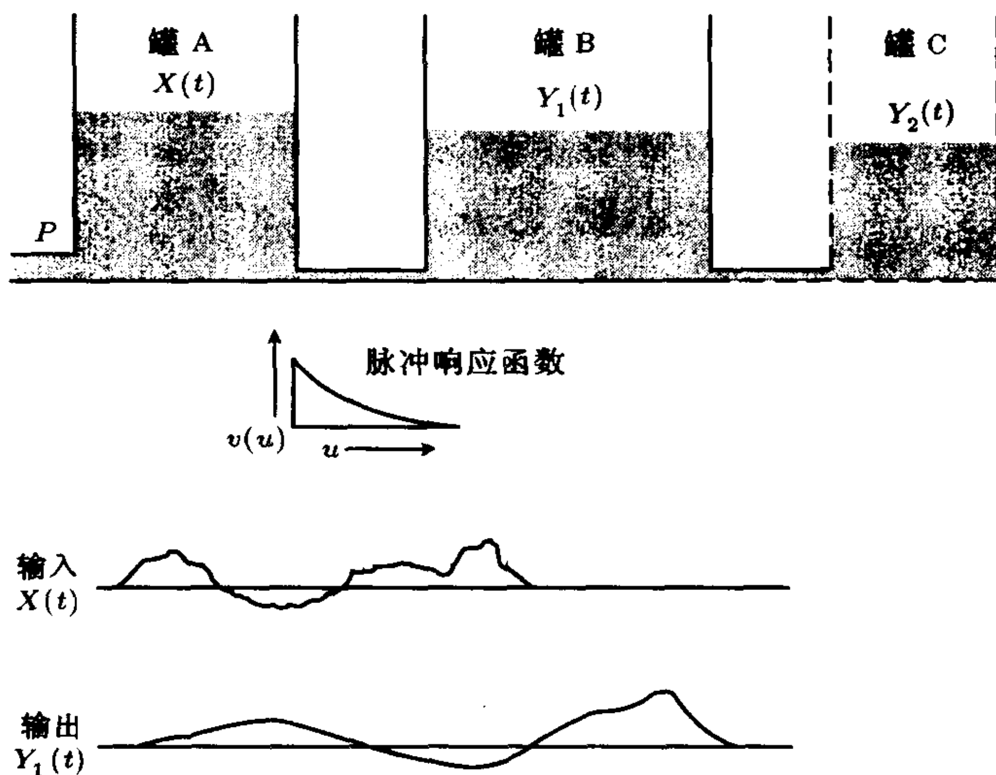


图 10.3　一个简单动态系统的表示

假设现在通过强制管道 P 中液体的输入和输出，使体积 $X(t)$ 遵从图 10.3 中“输入 $X(t)$”所标明的曲线形式，那么，B 中的体积 $Y_1(t)$ 会相应地按某种方式变化，在图中用 $Y_1(t)$ 标出了该曲线。一般地，$X(t)$ 对于驱动系统是决定性的，被称为强制函数。

为了把输入和输出联系起来，我们注意到在非常近似的情形下，管道中水流速度和前文的水平面差成比例. 即是

$$\frac{dY_1}{dt}=\frac{1}{T_1}[g_1X(t)-Y_1(t)] \tag{10.1.5}$$

这里 T_1 是常数. 微分方程 (10.1.5) 可以写成

$$(1+T_1D)Y_1(t)=g_1X(t) \tag{10.1.6}$$

其中 $D=d/dt$. 这种由一阶微分方程描述的动态系统常常称为一阶动态系统. 常数 $T-1$ 称为系统的时间常数. 同样的一阶模型可以近似地描述许多简单系统的特性. 例如，$Y_1(t)$ 可以是从一个加热器排出的水温，而 $X(t)$ 是进入加热器水的流速.

不难证明 (如，见 [122])，形如 (10.1.6) 的线性微分方程的解可以写为下面形式

$$Y_1(t)=\int_0^{\infty}v(u)X(t-u)du \tag{10.1.7}$$

这里 $v(u)$ 通常是 (连续) 的脉冲响应函数. 我们看到，$Y_1(t)$ 由 $X(t)$ 的连续加权组合生成，正如在 (10.1.2) 中，Y_t 由 X_t 的离散加权组合生成一样. 我们进一步看到，在连续情形 $v(u)$ 起着权函数的作用，这完全类似于在离散情形 v_j 所起的作用. 对于由 (10.1.6) 定义的具体的一阶系统有

$$v(u)=g_1T_1^{-1}e^{-u/T_1}$$

于是，在这种情形下如图 10.3 所示，脉冲响应函数服从简单的指数衰减。

在连续情形，对于完全任意的强制函数，其输出的确定通常是通过计算机的模拟来实现，或是利用在数字机上的数值过程来实现. 只是对特殊的强制函数其解析解可以得到. 例如，假定水压系统被置空，然后 $X(t)$ 突然升高到某一水平 $X(t)=1$, 并保持在这一水平上. 这种强制函数从稳定水平为 0 突然变化到稳定水平为 1, 我们称之为 (单位) 阶跃函数. 系统对这种函数的响应称为对系统的阶跃响应. 求解具有单位阶跃输入的微分方程，经推导可得

$$Y_1(t)=g_1(1-e^{-t/T_1}) \tag{10.1.8}$$

这样，罐 B 中水平面依照图 10.4 所示的指数形式增长. 现在，$t=T_1, Y_1(t)=g_1(1-e^{-1})=0.632g_1$. 因此，对于一阶系统 (10.1.6) 来说，从一个阶跃输入进入系统开始，直到达到最终均衡水平的 63.2%, 这一过程所需时间就是时间常数 T_1 .

有时对于一个给定的输入变化，其响应在开始起作用之前有一段纯延迟或停滞的时间. 例如，如果在图 10.3 中 A 和 B 有很长的管道， A 水平的突然变化要在液体流出管道之后才会起作用. 假定由此引入的延迟占用了 τ 个时间单位，那么延迟系统的响应就可以用类似于 (10.6) 的微分方程来描述，只不过把右边的 t 用 $t-\tau$ 来替换而已，所以

$$(1+T_1D)Y_1(t)=g_1X(t-\tau) \tag{10.1.9}$$

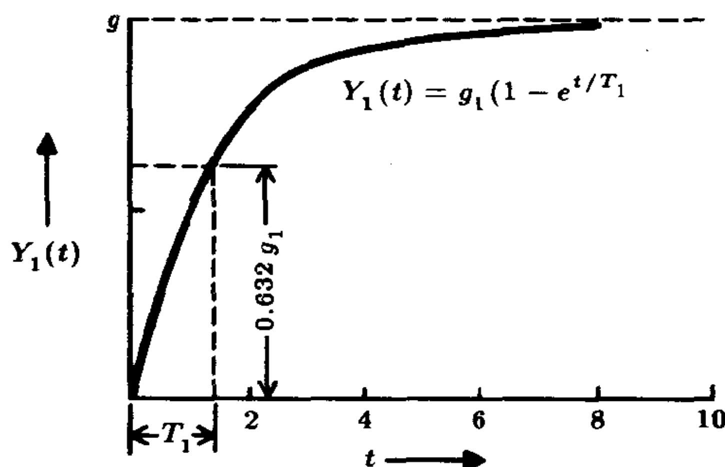


图 10.4 一阶系统对单位阶跃变化的响应

对于该系统而言，相应的脉冲和阶梯响应函数与无延迟系统完全相同，只不过是这些响应函数沿水平轴移动了距离 τ 。

二阶动态系统　再次考虑图 10.3. 设想一个三水罐的系统，有管道从罐 B 通向第三个水罐 C,C 中液体体积记为 $Y_2(t)$ 。令 T_2 为附加系统的时间常数， g_2 为稳态增益。那么 $Y_2(t)$ 和 $Y_1(t)$ 由下面微分方程相联系：

$$(1+T_2D)Y_2(t)=g_2Y_1(t)$$

把 (10.1.6) 带入，我们得到一个二阶微分方程，它把第三个罐的输出与第一个罐的输入联系起来

$$[1+(T_1+T_2)D+T_1T_2D^2]Y_2(t)=gX(t) \tag{10.1.10}$$

其中 $g=g_1g_2$. 这样一个系统的脉冲响应函数是指数混合型的

$$v(u)=\frac{g(e^{-u/T_1}-e^{-u/T_2})}{T_1-T_2} \tag{10.1.11}$$

对于单位阶跃的响应由下式给出

$$Y_2(t)=g(1-\frac{T_1e^{-t/T_1}-T_2e^{-t/T_2}}{T_1-T_2}) \tag{10.1.12}$$

图 10.5 中的连续曲线给出了系统

$$(1+3D+2D^2)Y_2(t)=5X(t)$$

对于单位阶跃的响应。这里 $T_1=1,T_2=2,g=5$ 。注意，与一阶系统不同的是，二阶系统的阶跃响应在一开始斜率为零。

更一般的二阶系统定义为

$$(1+\Xi_1D+\Xi_2D^2)Y(t)=gX(t) \tag{10.1.13}$$

其中

$$\Xi_1 = T_1 + T_2 \quad , \quad \Xi_2 = T_1 T_2 \tag{10.1.14}$$

而且常数 T_1 和 T_2 可以是复数。如果我们写成

$$T_1 = \frac{1}{\zeta} e^{i\lambda} \quad , \quad T_2 = \frac{1}{\zeta} e^{-i\lambda} \tag{10.1.15}$$

那么 (10.1.13) 成为

$$(1 + \frac{2\cos\lambda}{\zeta} D + \frac{1}{\zeta^2} D^2) Y(t) = gX(t) \tag{10.1.16}$$

脉冲响应函数化简成

$$v(u) = g \frac{\zeta e^{-\zeta u \cos\lambda} \sin(\zeta u \sin\lambda)}{\sin\lambda} \tag{10.1.17}$$

对于单位阶跃响应 (10.1.12) 为

$$Y(t) = g[1 - \frac{e^{-\zeta t \cos\lambda} \sin(\zeta t \sin\lambda + \lambda)}{\sin\lambda}] \tag{10.1.18}$$

图 10.5 中的连续曲线 C 给出了系统

$$(1 + \sqrt{2} D + 2D^2) Y(t) = 5X(t)$$

对于单位阶跃的响应，其中 $\lambda = \pi/3$ 及 $\zeta = \sqrt{2}/2$ 。我们注意到，响应在超过了值 $g = 5$ 之后，以一个阻尼正弦波趋于均衡，这是典型的欠阻尼系统的特征，其特征正如其名。通常，一个二阶系统称为过阻尼、临界阻尼或欠阻尼，取决于常数 T_1 和 T_2 是实数、相等实数或复数。过阻尼系统的阶跃响应是指数的混合型，由 (10.1.12) 给出，且始终保持在渐近线 $Y_\infty = g$ 之下。与一阶系统一样，把 (10.1.13) 右边的 t 换成 $t - \tau$，响应就可以反映出一段滞后时间。许多非常复杂的动态系统都可以用这种带延迟的二阶系统来很好地近似。

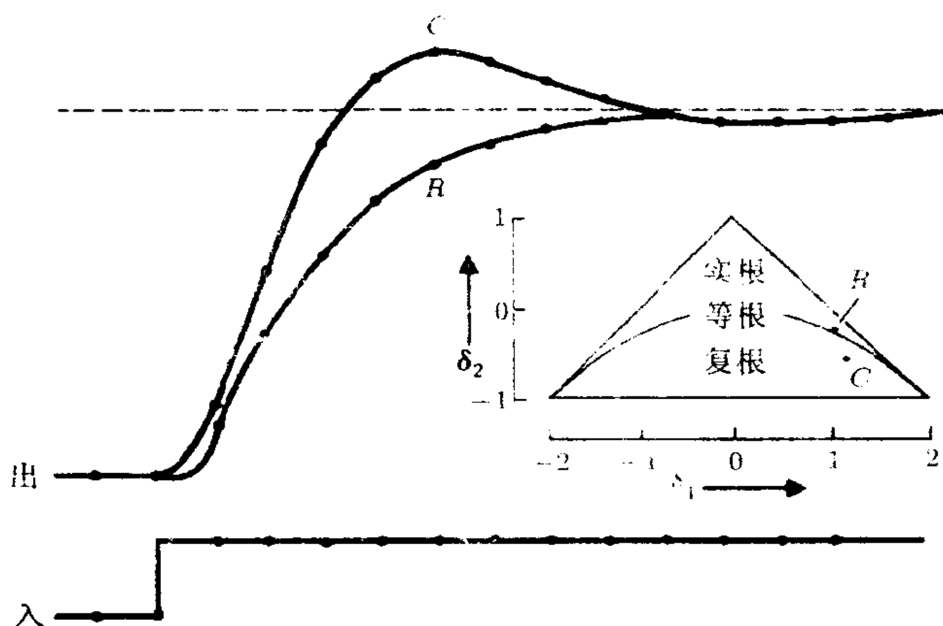


图 10.5 离散和连续二阶系统的同时阶跃响应，其特征方程具有实根（曲线 R）和复根（曲线 C）

更精细的线性动态系统可以描述为：不仅允许强制函数 $X(t)$ 的水平，而且允许其变化率 dX/dt 及高阶导数影响系统的特征。因此，描述（连续）动态系统的一般模型就是线性微分方程

$$(1+\Xi_1 D+\cdots\Xi_R D^R)Y(t)=g(1+H_1 D+\cdots+H_S D^S)X(t-\tau) \quad (10.1.19)$$

10.2 差分方程表示的离散动态模型

10.2.1 差分方程的一般形式

对应于连续形式 (10.1.19)，离散动态系统通常可简洁地用一般线性差分方程表示：

$$(1+\xi_1\nabla+\cdots+\xi_r\nabla^r)Y_t=g(1+\eta_1\nabla+\cdots+\eta_s\nabla^s)X_{t-b} \quad (10.2.1)$$

我们称之为 (r,s) 阶的传递函数模型。差分方程 (10.2.1) 还可以用后移算子 $B=1-\nabla$ 表示为

$$(1-\delta_1 B-\cdots-\delta_r B^r)Y_t=(\omega_0-\omega_1 B-\cdots-\omega_s B^s)X_{t-b} \quad (10.2.2)$$

或

$$\delta(B)Y_t = \omega(B)X_{t-b}$$

等价地，记 $\Omega(B) = \omega(B)B^b$, 则模型成为

$$\delta(B)Y_t = \Omega(B)X_t \tag{10.2.3}$$

比较 (10.2.3) 和 (10.1.2), 我们看到模型的传递函数为

$$v(B) = \delta^{-1}(B)\Omega(B) \tag{10.2.4}$$

因此，传递函数用两个 B 的多项式之比来表示。

ARIMA 模型的动态特性　ARIMA 模型

$$\varphi(B)z_t = \theta(B)a_t$$

用以描述时间序列 $\{z_t\}$, 还可以用线性滤波运算

$$z_t = \varphi^{-1}(B)\theta(B)a_t$$

来表示 z_t 和 a_t 之间的关系，其中 a_t 是白噪声。这样由 ARIMA 模型必然得知：时间序列可以表示为一个动态系统的输出，相应输入是白噪声，其传递函数可以简约地表示为 B 的两个多项式之比。

离散模型的稳定性　对于离散传递函数模型的稳定性要求与 ARMA 随机模型的平稳性要求完全类似. 为了稳定性，我们要求特征方程

$$\delta(B) = 0$$

的根在单位圆外，在上式中我们将 B 视为变量。特别地，对于一阶模型这蕴含参数 δ_1 满足

$$-1 < \delta_1 < 1$$

对于二阶模型 (例如见图 10.5), 参数 δ_1, δ_2 满足

$$\delta_2 + \delta_1 < 1 \quad , \quad \delta_2 - \delta_1 < 1 \quad , \quad -1 < \delta_2 < 1$$

将 (10.2.2) 详细地写出，有

$$Y_t = \delta_1 Y_{t-1} + \cdots + \delta_r Y_{t-r} + \omega_0 X_{t-b} - \omega_1 X_{t-b-1} - \cdots - w_s X_{t-b-s}$$

我们看到，若 X_t 恒取值 $-1, Y_t$ 最终将达到值

$$g = \frac{\omega_0 - \omega_1 - \cdots - \omega_s}{1 - \delta_1 - \cdots - \delta_r} \tag{10.2.5}$$

这是稳态增益用模型参数来表示的表达式。

10.2.2 传递函数的性质

如果我们考虑差分方程 (10.2.2) 定义的传递函数模型

$$Y_t = v(B) X_t \tag{10.2.6}$$

代入 (10.2.2), 我们得到恒等式

$$\begin{aligned}(1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r)(v_0 + v_1 B + v_2 B^2 + \cdots)\\ = (\omega_0 - \omega_1 B - \cdots \omega_s B^s) B^b\end{aligned} \tag{10.2.7}$$

令 B 的系数相等，我们得出

$$v_j = \begin{cases} 0 & , j < b \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots + \delta_r v_{j-r} + \omega_0 & , j = b \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots + \delta_r v_{j-r} + \omega_{j-b} & , j = b+1, b+2, \\ & \quad \cdots, b+s \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots + \delta_r v_{j-r} & , j > b+s \end{cases} \tag{10.2.8}$$

权数 $v_{b+s}, v_{b+s-1}, \cdots, v_{b+s-r+1}$ 为差分方程

$$\delta(B) v_j = 0 \qquad (j > b+s)$$

提供了 r 个初值。差分方程的解

$$v_j = f(\boldsymbol{\delta}, \boldsymbol{\omega}, j)$$

适用于 $j > b+s-r+1$ 的所有 v_j 。

因此，脉冲响应函数 v_j 的构成是：

1. b 个零值 $v_0, v_1, \cdots, v_{b-1}$ 。

2. 随后的 $s+r-1$ 个值 $v_b, v_{b+1}, \cdots, v_{b+s-r}$, 形式不固定 (当 $s<r$ 时，不出现这样的值)

3. 当 $j > b+s-r+1$ 时 ,v_j 的形式取决于 r 阶差分方程，其 r 个初值为 $v_{b+s}, v_{b+s-1}, \cdots, v_{b+s-r+1}$ 。当 $j<b$ 时，初值 v_j 当然是零。

阶跃响应　现在我们将 $V(B)$ 写成阶跃响应权数 V_j 的生成函数，V_j 表示在时刻 $j>0$ 对 0 时刻单位阶跃 (即 $X_t=1$, 当 $t>0; X_t=0$, 当 $t<0$) 的响应，因此，$V_j=\sum_{i=0}^{j} v_i$, 对 $j>0$ 。于是

$$\begin{aligned} V(B) &= V_0 + V_1 B + V_2 B^2 + \cdots \\ &= v_0 + (v_0+v_1)B + (v_0+v_1+v_2)B^2 + \cdots \end{aligned} \tag{10.2.9}$$

及

$$v(B) = (1-B)V(B) \tag{10.2.10}$$

将 (10.2.10) 代入 (10.2.7) 得到恒等式

$$\begin{aligned} &(1-\delta_1^* B - \delta_2^* B^2 - \cdots - \delta_{r+1}^* B^{r+1})(V_0 + V_1 B + V_2 B^2 + \cdots) \\ &= (\omega_0 - \omega_1 B - \cdots - \omega_s B^s) B^b \end{aligned} \tag{10.2.11}$$

这里

$$(1-\delta_1^* B - \delta_2^* B^2 - \cdots - \delta_{r+1}^* B^{r+1}) = (1-B)(1-\delta_1 B - \cdots - \delta_r B^r) \tag{10.2.12}$$

恒等式 (10.2.11) 是关于阶跃响应函数 V_j 的，它和脉冲响应函数的恒等式 (10.2.7) 完全相似，只是右边的算子 $\delta^*(B)$ 是 $r+1$ 阶而不是 r 阶。

利用 (10.2.8) 的结果，可以得出阶跃响应函数确定如下：

1. b 个零值 $V_0, V_1, \cdots, V_{b-1}$.

2. 随后的 $s-r$ 个值 $V_b, V_{b+1}, \cdots, V_{b+s-r-1}$, 形式不固定 (当 $s<r+1$ 时，不出现这样的值)

3. 当 $j > b+s-r$ 时，值 V_j 的形式取决于 $r+1$ 阶差分方程 $\delta^*(B)V_j=0$, 其 $r+1$ 个初值为 $V_{b+s}, V_{b+s-1}, \cdots, V_{b+s-r}$ 。当 $j<b$ 时，初值 V_j 当然是零.

10.2.3 一阶和二阶离散传递函数模型

表 10.1 给出了 $r=0,1,2$ 和 $s=0,1,2$ 所有组合的传递函数模型的详细情况. 图 10.6 给出了模型的特例，用棒形图表示其阶跃响应和脉冲响应. 表 10.1 最后的式子将 ∇ 形式下的模型参数 $\boldsymbol{\xi}, g, \boldsymbol{\eta}$ 用 B 形式下的参数 $\boldsymbol{\delta}, \boldsymbol{\omega}$ 表示出. 这些式子是由所考虑的最一般模型给出的，即 $r=2$ 和 $s=2$ 的模型. 所有其它模型都是这种情况的特例，令适当的参数为零可得到它们的相应表达式. 例如，若 $r=1, s=1$,$\xi_2=\eta_2=\delta_2=\omega_2=0$, 有

$$\delta=\frac{\xi_1}{1+\xi_1}, \qquad \omega_0=\frac{g(1+\eta_1)}{1+\xi_1}, \qquad \omega_1=\frac{g\eta_1}{1+\xi_1}$$

在表 10.2 中差分方程的初始值分别被脉冲和阶跃响应所满足，在棒形图中用小圆圈标记.

表中模型的讨论　在表 10.1 和图 10.6 中对一些模型的性质加以总结，我们将作详细地研究，因为这些模型描述经常遇到的动态系统是很有用的. 在所有模型中，右边的算子 B^b 使得脉冲响应函数的第一个非零项是 v_b . 图 10.6 的例子中， g 的值假设为 1, b 假设为 3.

r=0 的模型　r 和 s 都等于零时，脉冲响应只有一个值 $v_b=\omega_0=g$. 输出和输入成比例，但被移动了 b 个时间间隔. 更一般地，如果我们在右边有一个 s 阶算子，则瞬时输入将被延迟 b 个时间间隔，并按比例 $v_b=\omega_0, v_{b+1}=-\omega_1, \cdots, v_{b+s}=-\omega_s$ 分成

表 10.1　形如 $\delta_r(B)Y_t=\omega_s(B)B^bX_t$ 传递函数模型的脉冲响应函数

rsb	∇ 形式	B 形式	脉冲响应 v_j	
			0	$j<b$
00b	$Y_t=gX_{t-b}$	$Y_t=\omega_0B^bX_t$	ω_0	$j=b$
			0	$j>b$
			0	$j<b$
			ω_0	$j=b$
01b	$Y_t=g(1+\eta_1\nabla)X_{t-b}$	$Y_t=(\omega_0-\omega_1B)B^bX_t$	$-\omega_1$	$j=b+1$
			0	$j>b+1$
			0	$j<b$
			ω_0	$j=b$
02b	$Y_t=g(1+\eta_1\nabla+\eta_2\nabla^2)X_{t-b}$	$Y_t=(\omega_0-\omega_1B-\omega_2B^2)B^bX_t$	$-\omega_1$	$j=b+1$
			$-\omega_2$	$j=b+2$
			0	$j>b+2$
			0	$j<b$
10b	$(1+\xi_1\nabla)Y_t=gX_{t-b}$	$(1-\delta_1B)Y_t=\omega_0B^bX_t$	ω_0	$j=b$
			δ_1v_{j-1}	$j>b$
			0	$j<b$
			ω_0	$j=b$

注：$\xi_1=\frac{\delta_1+2\delta_2}{1-\delta_1-\delta_2}$，　$\xi_2=\frac{-2\delta_2}{1-\delta_1-\delta_2}$，　$\delta_1=\frac{\xi_1+2\xi_2}{1+\xi_1+\xi_2}$，　$\delta_2=\frac{-2\xi_2}{1+\xi_1+\xi_2}$，

$g=\frac{\omega_0-\omega_1-\omega_2}{1-\delta_1-\delta_2}$，　$\omega_0=\frac{g(1+\eta_1+\eta_2)}{1+\xi_1+\xi_2}$，　$\omega_1=\frac{g(\eta_1+2\eta_2)}{1+\xi_1+\xi_2}$，　$\omega_2=\frac{-g\eta_2}{1+\xi_1+\xi_2}$，

$\eta_1=\frac{\omega_1+2\omega_2}{\omega_0-\omega_1-\omega_2}$，　$\eta_2=\frac{-2\omega_2}{\omega_0-\omega_1-\omega_2}$，　$1-\delta_1\delta_2=(1+\xi_1+\xi_2)^{-1}$

表 10.1　形如 $\delta_r(B)Y_t=\omega_s(B)B^bX_t$ 传递函数模型的脉冲响应函数 (续)

rsb	∇ 形式	B 形式	脉冲响应 v_j	
$11b$	$(1+\xi_1\nabla)Y_t=g(1+\eta_1\nabla)X_{t-b}$	$(1-\delta_1B)Y_t=(\omega_0-\omega_1B)B^bX_t$	$\delta_1\omega_0-\omega_1$	$j=b+1$
			δ_1v_{j-1}	$j>b+1$
			0	$j<b$
			ω_0	$j=b$
$12b$	$(1+\xi_1\nabla)Y_t=g(1+\eta_1\nabla+\eta_1\nabla^2)X_{t-b}$	$(1-\delta_1B)Y_t=(\omega_0-\omega_1B-\omega_2B^2)B^bX_t$	$\delta_1\omega_0-\omega_1$	$j=b+1$
			$\delta_1^2\omega_0-\delta_1\omega_1-\omega_2$	$j=b+2$
			δ_1v_{j-1}	$j>b+2$
			0	$j<b$
$20b$	$(1+\xi_1\nabla+\xi_2\nabla^2)Y_t=gX_{t-b}$	$(1-\delta_1B-\delta_2B^2)Y_t=\omega_0B^bX_t$	ω_0	$j=b$
			$\delta_1v_{j-1}+\delta_2v_{j-2}$	$j>b$
			0	$j<b$
			ω_0	$j=b$
$21b$	$(1+\xi_1\nabla+\xi_2\nabla^2)Y_t=g(1+\eta_1\nabla)X_{t-b}$	$(1-\delta_1B-\delta_2B^2)Y_t=(\omega_0-\omega_1B)B^bX_t$	$\delta_1\omega_0-\omega_1$	$j=b+1$
			$\delta_1v_{j-1}+\delta_2v_{j-2}$	$j>b+1$
			0	$j<b$
			ω_0	$j=b$
$22b$	$(1+\xi_1\nabla+\xi_2\nabla^2)Y_t=g(1+\eta_1\nabla+\eta_2\nabla^2)X_{t-b}$	$(1-\delta_1B-\delta_2B^2)Y_t=(\omega_0-\omega_1B-\omega_2B^2)B^bX_t$	$\delta_1\omega_0-\omega_1$	$j=b+1$
			$(\delta_1^2+\delta_2)\omega_0$ $-\delta_1\omega_1-\omega_2$	$j=b+2$
			$\delta_1v_{j-1}+\delta_2v_{j-2}$	$j>b+2$

r,s,b	∇ 形式	B 形式	脉冲响应 v_j	阶跃响应 $V_j=\sum_{i=0}^{j} v_i$
003	$Y_t = X_{t-3}$	$Y_t = B^3 X_t$		
013	$Y_t=(1-.5\nabla)X_{t-3}$	$Y_t=(.5+.5B)B^3X_t$		
023	$Y_t=(1-\nabla+.25\nabla^2)X_{t-3}$	$Y_t=(.25+.50B+.25B^2)B^3X_t$		
103	$(1+\nabla)Y_t=X_{t-3}$	$(1-.5B)Y_t=.5B^3X_t$		
113	$(1+\nabla)Y_t=(1-.5\nabla)X_{t-3}$	$(1-.5B)Y_t=(.25+.25B)B^3X_t$		
123	$(1+\nabla)Y_t=(1-\nabla+.25^2)X_{t-3}$	$(1-.5B)Y_t=(.125+.25B+.125B^2)B^3X_t$		
203	$(1-.25\nabla+.5\nabla^2)Y_t=X_{t-3}$	$(1-.6B+AB^2)Y_t=.8B^3X_t$		
213	$(1-.25\nabla+.5\nabla^2)Y_t=(1-.5\nabla)X_{t-3}$	$(1-.6B+AB^2)Y_t=(.4+AB)B^3X$		
223	$(1-.25\nabla+.5\nabla^2)Y_t=(1-\nabla+.25^2)X_{t-3}$	$(1-.6B+AB^2)Y_t=(.2+AB+2B^2)B^3X_t$		

图 10.6　增益 $g=1$ 的脉冲响应和阶跃响应函数示例

$s+1$ 个值。阶跃响应可由脉冲响应求和得到，且最终满足初值为 $V_{b+s}=g=\omega_0-\omega_1-\cdots-\omega_s$ 的差分方程 $(1-B)V_j=0$ 。

r=1 的模型　$s=0$ 时，脉冲响应从初始值 $v_b=\omega_0=g/(1-\xi_1)=g(1-\delta_1)$ 开始，呈现指数（几何）拖尾。阶跃响应则指数增长直到使得值 $g=1$ 。如图中虚线所示，若指数阶跃响应向回推，则与时间轴交于时刻 $b-1$ 。这符合下面的事实：与 $V_b=v_b$ 一样，V_{b-1} 也是适当差分方程 $(1-\delta B)(1-B)V_j=0$ 的初始值。

$s=1$ 时，脉冲响应的一个初始值是 $v_b=\omega_0=g(1+\eta_1)/(1+\xi_1)$，但它不服指数形式。与方程左边算子相联系的差分方程 $v_j=\delta_1 v_{j-1}$ 可以推出指数形式初值，并以 $v_{b+1}=(\delta_1\omega_0-\omega_1)=g(\xi_1-\eta_1)/(1-\xi_1)^2$ 为起始值。阶跃响应函数服从于由差分方程 $(1-\delta B)(1-B)V_j=0$ 所决定的指数曲线，它由初值 $V_b=v_b$ 和 $V_{b+1}=v_b+v_{b+1}$ 开始渐近趋于 g 。在图中用虚线向后连接各点所画出的指数曲线，通常与时间轴相交于时间间隔之中的某点。在节 10.3 中我们将说明，对于近似具有小数时间延迟的连续一阶系统，实际上可以用一阶差分方程来描述，其右边为阶数 $s=1$ 的算子。

$s=2$ 时，脉冲响应有两个值 v_b 和 v_{b+1} 不服从指数衰减，从 v_{b+2} 开始服从指数衰减。相应地，在阶跃响应中，有一个初值 V_b 与用虚线连接的指数曲线不一致。如前所述，该曲线由差分方程 $(1-\delta B)(1-B)V_j=0$ 所决定，但初始值为 V_{b+1} 和 V_{b+2} 。

r=2 的模型　$s=0$ 时模型的灵活性是有限的，因为脉冲响应的第一个初始值固定为零。对 $s=1$ 和 $s=2$ 可得到更有用的模型。这些模型用来近似连续的二阶系统，我们将在节 10.3 和附录 A10.1 中加以讨论。

动态权数 v_j 最终满足

$$v_j-\delta_1 v_{j-1}-\delta_2 v_{j-2}=0 \qquad (j>b+s) \tag{10.2.13}$$

其特性依赖于特征方程

$$1-\delta_1 B-\delta_2 B^2=(1-S_1 B)(1-S_2 B)=0$$

表 10.2 二阶系统对于 $1-\delta_1 B-\delta B^2$ 根的依赖关系

根 (S_1^{-1},S_2^{-1})	条件	阻尼
实根	$\delta_1^2+4\delta_2>0$	过阻尼
相等实根	$\delta_1^2+4\delta_2=0$	邻界阻尼
复根	$\delta_1^2+4\delta_2<0$	欠阻尼

根 S_1^{-1} 和 S_2^{-1} 的性质。这种依赖关系在表 10.2 中给出。和连续情形一样，模型可以是过阻尼、邻界阻尼或欠阻尼的，这取决于特征方程根的性质。

在复根情形，(10.2.13) 的解服从阻尼正弦振荡，在图 10.6 中给出了二阶系统的示例。当根为实根时，其解是两个拖尾指数之和。正如在节 10.1.2 中所考虑的连续情形一样，该系统可以被认为等价于由原序列分出的两个离散的一阶系统，具有的参数为 S_1 和 S_2 。

阶跃响应的权数 V_j 最终满足差分方程

$$(V_j-g)-\delta_1(V_{j-1}-g)-\delta_2(V_{j-2}-g)=0$$

这与 (10.2.13) 形式相同。因此，阶跃响应在其渐近值 g 的变化特性类似于脉冲响应在时间轴上下的变化特性。在复根的情形，阶跃响应先是超越 g 的值，然后围绕该值波动，直至达到均衡。当根是正实根时，阶跃响应是两个指数项之和，它趋于渐近值 g, 但不会穿越该值。然而，如果有负实根，阶跃响应就会超越并波动直到最后稳定。当根是正实根时，阶跃响应是两个指数项之和，它趋于渐近值 g, 但不会穿越该值。然而，如果有负实根，阶跃响应就会超越并波动直到最后稳定在均衡值。

在图 10.5 中的点表示两个离散的阶跃响应，分别用 R 和 C 标出，这些点对应于图形底部的点所表示的离散阶跃输入。相应于 R 和 C 的差分方程[①] 是:

---

① 事实上，这些模型的参数是用节 10.3.2 中讨论的方法选择的，因此，在离散点上，阶跃响应与节 10.1.2 中引入的连续系统完全匹配。

$$\begin{aligned}
\text{R}: &\quad (1-0.97B+0.22B^2)Y_t = 5(0.15+0.09B)X_{t-1} \\
\text{C}: &\quad (1-1.15B+0.49B^2)Y_t = 5(0.19+0.15B)X_{t-1}
\end{aligned}$$

图 10.5 中还给出了稳定域的图，并标出了两个模型的参数点对 (δ_1,δ_2) 。注意，采用模型 R 的系统具有不穿越的实根，而模型 C 具有复根，并发生穿越。

10.2.4 对于任意输入递推计算输出

如果我们必须使用模型的脉冲响应形式 (10.1.2) 对于给定输入去计算输出，那将是非常繁琐的。所幸不必如此。另外，我们可以直接利用差分方程。不论对于什么样的输入，用这种方式递推地计算输出都是很简单的。例如，考虑 $r=1, s=0$, 及 $\xi=1, g=5$ 的模型，便有

$$(1+\nabla)Y_t = 5X_{t-1}$$

或等价地

$$(1-0.5B)Y_t = 2.5X_{t-1} \tag{10.2.14}$$

表 10.3 给出了 Y_t 的计算，输入 X_t 为： (a) 单位脉冲输入； (b) 单位阶跃输入； (c)“一般 ” 输入。在所有情形都假设输入具有初值 $Y_0=0$ 。为了进行递推计算，把 Y_t 写在差分方程的左边，即

$$Y_t = 0.5Y_{t-1} + 2.5X_{t-1}$$

例如在“一般 ” 输入的情形，

$$\begin{aligned}
Y_1 &= 0.5\times 0 + 2.5\times 0 = 0 \\
Y_2 &= 0.5\times 0 + 2.5\times 1.5 = 3.75
\end{aligned}$$

等等。这些输入和输出都在图 10.7 的 (a),(b) 和 (c) 中画出。

一般地，我们看到，只要把传递函数模型写成如下形式

$$Y_t = \delta_1 Y_{t-1} + \cdots + \delta_r Y_{t-r} + \omega_0 X_{t-b} - \omega_1 X_{t-b-1} - \cdots - \omega_s X_{t-b-s}$$

表 10.3　离散一阶系统对于脉冲，阶跃和一般输入的输出计算

	(a) 脉冲输入		(b) 阶跃输入		(c) 一般输入	
	输入	输出	输入	输出	输入	输出
t	X_t	Y_t	X_t	Y_t	X_t	Y_t
0	0	0	0	0	0	0
1	1	0	1	0	1.5	0
2	0	2.50	1	2.50	0.5	3.75
3	0	1.25	1	3.75	2.0	3.12
4	0	0.62	1	4.38	1.0	6.56
5	0	0.31	1	4.69	–2.5	5.78
6	0	0.16	1	4.84	0.5	–3.36

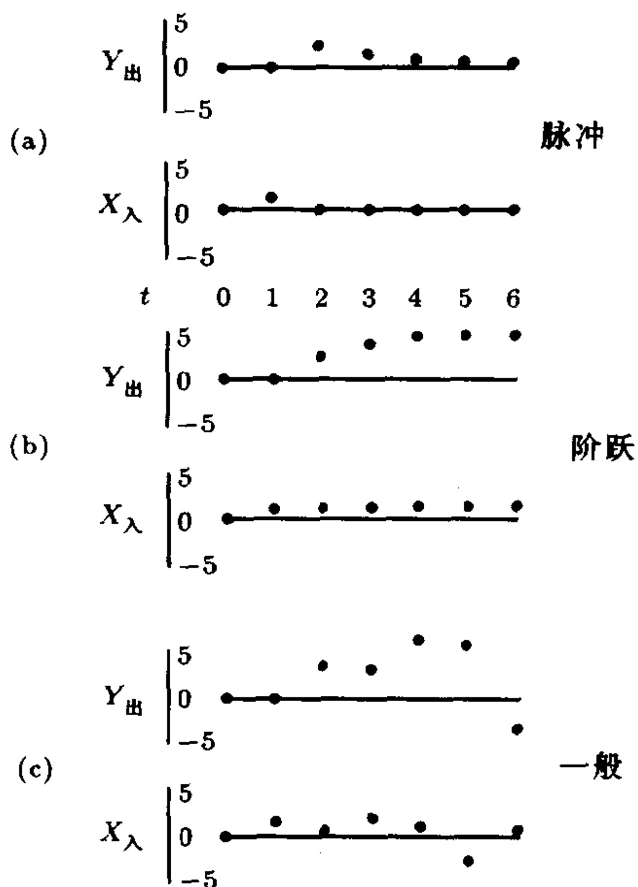


图 10.7　一阶系统对于脉冲，阶跃和"一般"输入的响应

对于任意离散输入去计算离散输出就是很简单的事了. 为了开始进行递推, 我们需要知道某些初始值. 这一要求当然不是这种计算方法的缺点, 但毕竟总需要去做, 而在传递函数模型中 Y 的初值依赖于开始观测之前出现的 X 值. 在实际中, 当必要的初值未知时, 我们可用均值来代替未知的 Y 和 X(如果这些量考虑为与均值的离差, 则用零来代替) . 较靠前的计算值将依赖于初值的选取, 但对于稳定系统而言, 在足够长的一段时间之后, 脉冲响应成为可以忽略的, 那么, 初值选取的影响也就可以忽略了. 若这段时间是 p_0 个时间间隔, 另一种处理方法是: 直接从脉冲响应计算 $Y_{p_0}, Y_{p_0+1}, \cdots$, 直到获得的值足以着手进行递推计算.

10.2.5 附加噪声的传递函数模型

在实际中, 不能指望输出严格遵从传递函数所确定的形式, 即便模型完全恰当也是如此. X 之外的各种干扰通常都会影响系统. 一种干扰可能源于系统中的任一点, 但为了简便起见, 如图 1.5 中所示, 我们考虑单纯只对输出 Y 的影响. 若我们假设干扰或噪声 N_t 与 X_t 的水平独立, 且添加到有关 X 的影响上, 我们写成

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t \tag{10.2.15}$$

如果噪声模型可以用 ARIMA(p,d,q) 过程来表示

$$N_t = \varphi^{-1}(B)\theta(B)a_t$$

这里, a_t 是白噪声, 模型 (10.2.15) 最终可写为

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + \varphi^{-1}(B)\theta(B)a_t \tag{10.2.16}$$

在第 11 章中, 我们要讲述形如 (10.2.16) 的传递函数 - 噪声混合模型的识别、拟合及检验.

10.3 离散模型和连续模型的关系 [①]

由线性差分方程定义的离散动态模型本身是非常重要的。它提供了一类切实可行的传递函数并且无需其它调整。是否可以将离散模型和假设的基本连续模型联系起来，在许多例子中并没有提出这样的问题，这是因为基本的连续序列本身就不存在。然而在有些情况下，例如，对一个化学反应器定时地采集瞬时观测值，离散记录可以告诉我们关于连续系统的一些情况。特别地，控制工程师利用这些记录，并据此考虑连续系统的时间常数和停滞时间，这样才能更好地理解离散分析的结果。

如前所述，我们分别用 $Y(t)$ 和 $X(t)$ 记时刻刻 t 的连续输出和输入。假设输出和输入用线性滤波运算相联系

$$Y(t) = \int_0^\infty v(u)X(t-u)du$$

现在假设只有输出和输入的离散观测值 $(X_t, Y_t), (X_{t-1}, Y_{t-1}), \cdots$，它们是在等间隔的时间 $t, t-1, \cdots$ 得到的，把离散的输出和输入联系在一起的是离散线性滤波器：

$$Y_t = \sum_{j=0}^{\infty} v_j X_{t-j}$$

那么，在某些特定情况和适当的假设下，离散模型和连续模型之间可建立起有益的关系。

10.3.1 对于脉冲化输入的响应

在一些特殊场合，输出值被观测到后立刻就有机会对过程进行调节，使得输入变量在两次观测之间保持同一水平，这种情形对于第 4 部分讨论的离散控制方案设计十分重要。在图 10.8 中

[①] 10.3 节的内容在初次阅读时可以略去。

给出了合成方波的典型图形，我们将称之为脉冲化输入。我们把在时间 $t-1<\tau<t$ 期间输入所保持的固定水平记为 X_{t-1+} 。

考虑一个连续线性系统，它有 b 个整时段的延迟再加上一个分数时段 c 的另外的延迟。于是用前面的符号来记就是 $b+c=\tau$ 。因此，我们可将系统的输出表示为

$$Y(t)=\int_0^\infty v(u)X(t-u)du$$

其中，脉冲响应函数 $v(u)$ 在 $u<b+c$ 时是零。现在，对于如图 10.9 中所示的脉冲化的输入， t 时刻的输出将由下式精确给出

$$Y(t)=\left[\int_{b+c}^{b+1} v(u)du\right]X_{t-b-1+}+\left[\int_{b+1}^{b+2} v(u)du\right]X_{t-b-2+}+\cdots$$

从而

$$Y(t)=Y_t=v_bX_{t-b-1+}+v_{b+1}X_{t-b-2+}+\cdots$$

因此，对于一个脉冲化输入，存在一个离散线性滤波器，使得在时刻 $t,t-1,t-2,\cdots$ 连续输出 $Y(t)$ 精确地等于离散输出。

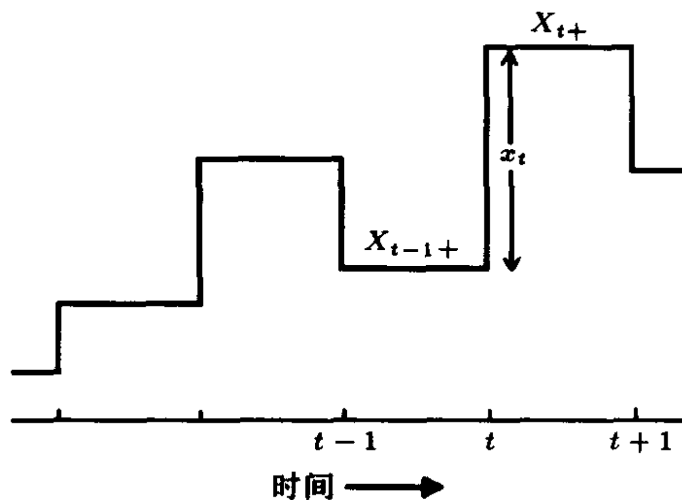


图 10.8 脉冲化输入示例

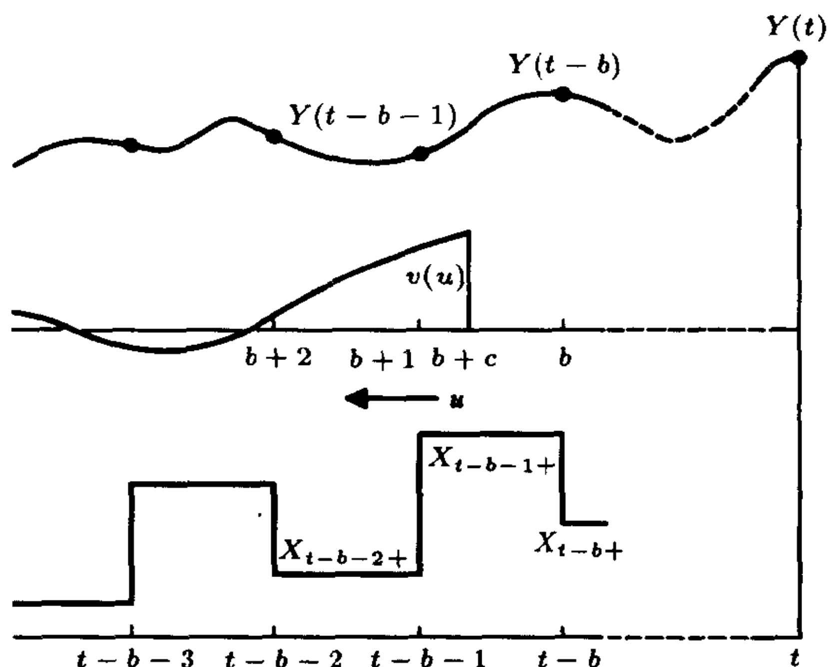


图 10.9 脉冲化输入到输出的传递

给定一个脉冲化输入，输出 $Y(t)$ 考虑为来自下面的 (r,r) 阶模型：

$$\xi(\nabla)Y_t = \eta(\nabla)X_{t-b-1+} \tag{10.3.1}$$

再考虑具有连续输出的 R 阶模型：

$$(1+\Xi_1 D+\Xi_2 D^2+\cdots+\Xi_E D^R)Y(t) = X(t-b-c) \tag{10.3.2}$$

这里，和前一模型有着同样的输入。附录 A10.1 证明了：适当选取参数 (Ξ,c) 的数值，当 $R=r$ 时两个输出将完全重合。进而，如果 $c=0$，连续模型 (10.3.2) 的输出会在离散时刻和 $(r,r-1)$ 阶离散模型 (10.3.1) 的输出完全相等。我们把这种相互关联的连续模型和离散模型称为离散相合系统。那么，如果 (r,r) 阶形如 (10.3.1) 的离散模型已经得到，按照假设，连续模型可由 r 阶微分方程 (10.3.2) 来表示，其参数 —特别是离散相合的连续系统的时间常数 —可以用离散模型的参数精确地表示。

在附录 A10.1 中对有延迟的二阶系统推导了参数关系。由此，对于较简单系统的相应关系，通过将适当的参数设为零就可以得到，在下文我们马上就要讨论有关内容.

10.3.2 一阶和二阶相合系统的关系

无延迟的一阶系统

B 形式　连续系统满足

$$(1+TD)Y(t)=gX(t) \tag{10.3.3}$$

离散系统满足

$$(1-\delta B)Y_t=\omega_0 X_{t-1+} \tag{10.3.4}$$

对于脉冲化输入，这两个系统是离散相合的，且有

$$\delta=e^{-1/T}\ , \qquad T=(-\ln\delta)^{-1}\ , \qquad \omega_0=g(1-\delta) \tag{10.3.5}$$

∇ 形式　另一方面，差分方程可以写成

$$(1+\xi\nabla)Y_t=gX_{t-1+} \tag{10.3.6}$$

其中

$$\xi=\frac{\delta}{1-\delta} \tag{10.3.7}$$

为了说明问题，我们考虑节 10.2.4 中"一般"输入的例子．这种情形的输出已在表 10.3(c) 中计算并在图 10.7(c) 中绘出．事实上，假设我们有一个连续系统

$$(1+1.44D)Y(t)=5X(t)$$

那么，它将与离散模型 (10.2.14) 是离散相合的，实际模型即

$$(1-0.5B)Y_t=2.5X_{t-1+}$$

如果输入和输出皆为连续，且输入是脉冲化的，响应的实际过程在图 10.10 中用连贯的实线给出．事实上，输出将遵从一连串的指数曲线．每段虚线所表示的是：当输入没有进一步变化时，响应将采取的发展过程．这些曲线和具有离散输出的离散样本点完全对应，离散情形在表 10.3(c) 中计算并在图 10.7(c) 绘出．

有延迟的一阶系统

B 形式　连续系统满足

$$(1+TD)Y(t)=gX(t-b-c) \tag{10.3.8}$$

对于脉冲化输入，它离散相合于满足下式的离散系统

$$(1-\delta B)Y_t=(\omega_0-\omega_1 B)X_{t-b-1+} \tag{10.3.9}$$

这里

$$\delta=e^{-1/T}\quad ,\ \omega_0=g(1-\delta^{1-c})\quad ,\ \omega_1=g(\delta-\delta^{1-c}) \tag{10.3.10}$$

∇ 形式　另一方面，差分方程可以写成

$$(1+\xi\nabla)Y_t=g(1+\eta\nabla)X_{t-b-1} \tag{10.3.11}$$

其中

$$\xi=\frac{\delta}{1-\delta}\quad ,\quad -\eta=\frac{\delta(\delta^{-c}-1)}{1-\delta} \tag{10.3.12}$$

现在有

$$(1+\eta\nabla)X_{t-b-1+}=(1+\eta)X_{t-b-1+}-\eta X_{t-b-2+} \tag{10.3.13}$$

该式可以看作是 X_{t-b-1+} 和 X_{t-b-2+} 以 η 为增量的插值。表 10.4 中列出了离散和连续模型的对应参数 $(\xi,-\eta)$ 和 (T,c)，以及其它范围所确定的模型。

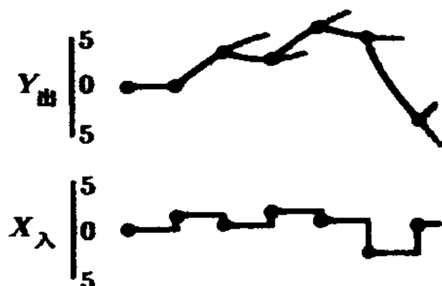


图 10.10　系统 $(1+1.44D)Y(t)=5X(t)$ 对脉冲化输入的连续响应

表 10.4　一阶有延迟系统与 T 和 c 各种值相对应的 $-\eta$ 值；对应的 ξ 和 δ 值

δ	ξ	T	$-\eta$ 关于				
			c=0.9	c=0.7	c=0.5	c=0.3	c=0.1
0.9	9.00	9.49	0.90	0.69	0.49	0.29	0.10
0.8	4.00	4.48	0.89	0.68	0.47	0.28	0.09
0.7	2.33	2.80	0.88	0.66	0.46	0.26	0.09
0.6	1.50	1.95	0.88	0.64	0.44	0.25	0.08
0.5	1.00	1.44	0.87	0.62	0.41	0.23	0.07
0.4	0.67	1.09	0.85	0.60	0.39	0.21	0.06
0.3	0.43	0.83	0.84	0.57	0.35	0.19	0.05
0.2	0.25	0.62	0.82	0.52	0.31	0.15	0.04
0.1	0.11	0.43	0.77	0.45	0.24	0.11	0.03

无延迟的二阶系统

B 形式　连续系统满足

$$(1+T_1D)(1+T_2D)Y(t)=gX(t) \tag{10.3.14}$$

对于脉冲化输入，它与下面系统离散相合：

$$(1-\delta_1B-\delta_2B^2)Y_t=(\omega_0-\omega_1B)X_{t-1+} \tag{10.3.15}$$

或等价地，系统为

$$(1-S_1B)(1-S_2B)Y_t=(\omega_0-\omega_1B)X_{t-1+} \tag{10.3.16}$$

这里

$$\begin{aligned} &S_1=e^{-1/T_1}\quad,\quad S_2=e^{-1/T_2}\\ &\omega_0=g(T_1-T_2)^{-1}[T_1(1-S_1)-T_2(1-S_2)]\\ &\omega_1=g(T_1-T_2)^{-1}[T_1S_2(1-S_1)-T_2S_1(1-S_2)] \end{aligned} \tag{10.3.17}$$

∇ 形式　另一方面，差分方程可以写成

$$(1+\xi_1\nabla+\xi_2\nabla^2)Y_t=g(1+\eta_1\nabla)X_{t-1+} \tag{10.3.18}$$

这里

$$-\eta_1=(1-S_1)^{-1}(1-S_2)^{-1}(T_1-T_2)^{-1}[T_2S_1(1-S_2)-T_1S_2(1-S_1)] \tag{10.3.19}$$

它可认为是 X_{t-1+} 和 X_{t-2+} 之间插值的增量。用 δ 表示的 ξ_1 和 ξ_2 值可直接用表 10.1 给出的结果得到。

作为特例，图 10.5 给出了前面考虑的两个离散系统的阶跃响应，它们都具有从离散相合系统得出相应的连续响应。

对于曲线 C, 这对模型是：

连续： $(1+1.41D+2D^2)Y(t)=5X(t)$

离散： $(1-1.15B+0.49B^2)Y_t=5(0.19+0.15B)X_{t-1+}$

对于曲线 R 有

连续： $(1+2D)(1+D)Y(t)=5X(t)$

离散： $(1-0.97B+0.22B^2)Y_t=5(0.15+0.09B)X_{t-1+}$

连续曲线是利用 (10.1.18) 和 (10.1.12) 画出的，它们分别给出了具有复根和实根的二阶系统的连续阶跃响应。

10.3.3 用离散模型近似一般的连续模型

或许我们应该再次强调，离散的传递函数模型不必用连续系统加以说明，或与之相联系. 可以从离散输入计算出离散输出，这一点本身就是很重要的。但是，在一些场合，二者之间的关系是有意义的.

对于连续系统来说，脉冲式输入本身是由控制问题引出的，一种简便的操作方法是：在取得输出 Y 的观测之后，立刻对输入变量 X 做一些需要的调整。这样输入变量停留在在两次观测值之间的一个固定水平上，我们就有了一个脉冲式的输入。前一节

建立起来的关系可以马上应用了。特别地，这些关系表明：按照我们的符号，无延迟离散系统可表示为

$$\xi(\nabla)Y_t = \eta(\nabla)X_{t-1+}$$

其中， X 的下标 $t-1+$ 比 Y 的下标 t 落后一步。

当连续记录得到时对离散模型的运用　正如我们在第 11 章中所述，尽管我们已经有了输入和输出的连续记录，用离散方法去确定系统的动态特性仍然是很方便的。如果用足够短的采样间隔来读取成对值，那么用离散记录来代替连续记录就几乎不会有什么损失。用离散结果来近似连续传递函数的一种方法是把输入当成脉冲式来处理，即如果离散输入是在时刻 t 观测到的，我们把输入记录处理为从 $j-\frac{1}{2}$ 拓展到 $j+\frac{1}{2}$，如图 10.11 中所示。因此有 $X(t)=X_j(j-\frac{1}{2}<t<j+\frac{1}{2})$．利用脉冲输入化方程，并用 X_t 代替 X_{t+}，用 $b+c-\frac{1}{2}$ 代替 $b+c$，即从延迟中减去半个时间单位，这样我们就可以把离散结果和连续结过联系起来。

对于连续记录通常是在足够小的采样间隔上读取，因此在样本点之间不会发生突变，在这种情形下，近似就是很准确的了。

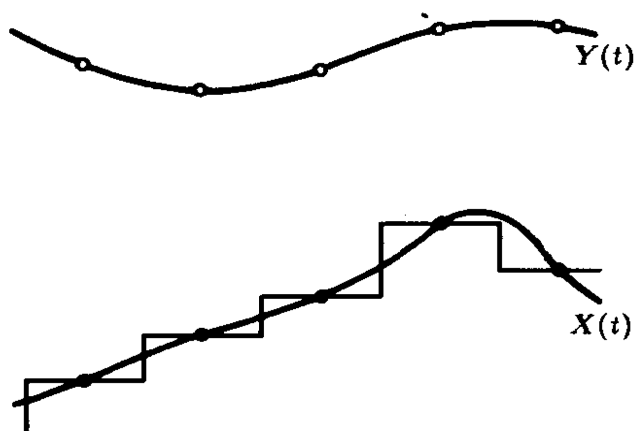


图 10.11　用脉冲式输入代替连续输入

A10.1 具有脉冲式输入的连续模型

在节 10.3.1(又可见图 10.9) 中我们说明了, 对于脉冲式输入, 任何有延迟连续线性系统的输出是

$$Y(t) = \int_0^\infty v(u)X(t-u)du$$

其中 $v(u) = 0, u < b+c$, 它在离散时刻 $t, t-1, t-2, \cdots$ 可以由离散线性滤波器

$$Y_t = v(B)X_{t-1+}$$

准确地给出。这里 $v_0, v_1, \cdots, b_{b-1}$ 为零，权数 $v_b, v_{b-1}, \cdots$ 由下式给出

$$v_b = \int_{b+c}^{b+1} v(u)du \tag{A10.1.1}$$

$$v_{b+j} = \int_{b+j}^{b+j+1} v(u)du \qquad (j > 1) \tag{A10.1.2}$$

现在假定连续系统的动态特征由 R 阶微分方程

$$\Xi(D)Y(t) = gX(t-b-c) \tag{A10.1.3}$$

描述，上式可写为如下形式

$$\prod_{h=1}^{R}(1+T_hD)Y(t) = gX(t-b-c)$$

这里， $T_1, T_2, \cdots, T_R$ 可以是实的或是复的。我们现在来说明，对于一个脉冲式输入，上述连续系统的输出离散地相合于一个离散差分方程模型的输出，模型的阶为 (r, r), 或当 $c = 0$ 时，阶为 $(r, r-1)$ 。这里，对于 $u < b+c, v(u)$ 为零，而对于 $u > b+c, v(u)$ 一般为非零. 且满足下面微分方程

$$\prod_{h=1}^{R}(1+T_hD)v(u-b-c) = 0 \qquad (u > b+c)$$

于是

$$v(u) = 0 \qquad (u < b + c)$$
$$v(u) = \alpha_1 e^{-(u-b-c)/T_1} + \alpha_2 e^{-(u-b-c)/T_2} + \cdots + \alpha_R e^{-(u-b-c)/T_R} \qquad (u \geqslant b + c)$$

因此，利用 (A10.1.1) 和 (A10.1.2) 有

$$v_b = \sum_{h=1}^{R} \alpha_h T_h [1 - e^{-(1-c)/T_h}] \tag{A10.1.4}$$

$$v_{b+j} = \sum_{h=1}^{R} \alpha_h T_h (1 - e^{-1/T_h}) e^{c/T_h} e^{-j/T_h} \qquad (j \geqslant 1) \tag{A10.1.5}$$

我们注意到，在 $c = 0$ 的特殊情形，当 $j = 0$ 和 $j > 0$ 时，权数 v_{b+j} 都由 (A10.1.2) 给出。

现在来考虑 (r, s) 阶差分方程模型

$$\delta(B) Y_t = \omega(B) B^b X_{t-1+} \tag{A10.1.6}$$

如果我们记

$$\Omega(B) = \omega(B) B^b$$

则该模型的离散传递函数 $v(B)$ 满足

$$\delta(B) v(B) = \Omega(B) \tag{A10.1.7}$$

正如我们在 (10.2.8) 中观察到的，通过令 (A10.1.7) 的系数相等，我们得到 b 个零值的权数 $v_o, v_1, \cdots, v_{b-1}$，当 $s \geqslant r$ 时，后面的 $s - r + 1$ 个权值 $v_b, v_{b+1}, \cdots, v_{b+s-r}$ 不遵从某一形式。权数 v_j 最终满足

$$\delta(B) v_j = 0 \qquad (j > b + s) \tag{A10.1.8}$$

$v_{b+s}, v_{b+s-1}, \cdots, v_{b+s-r}$ 给出所需要的 r 个初值．现在记

$$\delta(B) = \prod_{h=1}^{r} (1 - S_h B)$$

这里 $S_1^{-1}, S_2^{-1}, \cdots, S_r^{-1}$ 是方程 $\delta(B)=0$ 的根，则 (A10.1.8) 的解形式为

$$v_j = A_1(\boldsymbol{\omega})S_1^j + A_2(\boldsymbol{\omega})S_2^j + \cdots + A_r(\boldsymbol{\omega})S_r^j \qquad (j > b+s-r) \tag{A10.1.9}$$

其中系数 $A_h(\boldsymbol{\omega})$ 适当选取，以使 (A10.1.9) 的解当 $j = s-r+1, s-r+2, \cdots, s$ 时，能够得出初值 $v_{b+s-r+1}, \cdots, v_{b+s}$，使用符号 $A_h(\boldsymbol{\omega})$ 是为了提醒我们： A 是 $\omega_0, \omega_1, \cdots, \omega_s$ 的函数。因此，如果我们令 $s=r$，对于 (A10,1.6) 中给定的参数 $(\boldsymbol{\omega}, \boldsymbol{\delta})$— 也就是对于给定的参数 $(\boldsymbol{\omega}, \boldsymbol{S})$，将会有一组对应的值 $A_h(\boldsymbol{\omega})(h=1,2,\cdots,r)$ 来生成 r 个适当的初值 $v_{b+1}, v_{b+2}, \cdots, v_{b+r}$ 。此外，我们已知 $v_b = \omega_0$，因此

$$v_b = \omega_0 \tag{A10.1.10}$$

$$v_{b+j} = \sum_{h=1}^{r} A_h(\boldsymbol{\omega})S_h^j \tag{A10.1.11}$$

(A10.1.4) 和 (A10.1.11) 中的权来自于微分方程，而 (A10.1.10) 和 (A10.1.11) 中的权来自于差分方程，我们可以令这些权值相等。为此必须设

$$R = r \quad , \qquad S_h = e^{-1/T_h}$$

余下的 $r+1$ 个方程

$$\omega_0 = \sum_{h=1}^{r} \alpha_h T_h (1 - S_h^{1-c})$$

$$A_h(\boldsymbol{\omega}) = \alpha_h T_h (1 - S_h) S_h^{-c}$$

用来确定由 S 和 ω 表示的 $c, \alpha_1, \alpha_2, \cdots, \alpha_r$ 。

当 $c=0$ 时，我们令 $s=r-1$，为了给出差分方程中的参数 $(\boldsymbol{\omega}, \boldsymbol{S})$，将会用到一组 r 个值 $A_h(\boldsymbol{\omega})$，它们是 $\omega_0, \omega_1, \cdots, \omega_{r-1}$ 的函数，由 r 个初值 $v_b, b_{b+1}, \cdots, v_{b+r-1}$ 产生，且等于由 (A10.1.5) 式对 $j=0,1,\cdots,r-1$ 给出的值。为此我们令

$$R = r \quad , \qquad S_h = e^{-1/T_h}$$

余下的 r 个方程

$$A_h(\omega) = \alpha_h T_h(1 - S_h)$$

用来确定由 S 和 ω 表示的 $\alpha_1, \alpha_2, \cdots, \alpha_r$. 所以一般来说，对于脉冲式输入，由

$$\Xi(D)Y(t) = gX(t-b-c) \tag{A10.1.12}$$

定义的连续 r 阶动态系统在时刻 $t, t-1, \cdots$ 的输出，等同于如下离散模型的输出：

$$\xi(\nabla)Y_t = g\eta(\nabla)X_{t-b-1+} \tag{A10.1.13}$$

该模型为 (r, r) 阶，其参数适当选定. 此外，若 $c = 0$, 连续模型 (A10.1.12) 的输出在离散的时间点上等同于 $(r, r-1)$ 阶模型 (A10.1.13) 的输出。

我们现在来推导对应于二阶有延迟系统的离散模型，由此，在节 10.3.2 中给出的结果将作为特例而得出。

二阶有延迟系统　假设联系一个连续系统输入和输出的微分方程由下式给出

$$(1 + T_1 D)(1 + T_2 D)Y(t) = gX(t-b-c) \tag{A10.1.14}$$

那么，连续的脉冲响应函数为

$$v(u) = g(T_1 - T_2)^{-1}(e^{-(u-b-c)/T_1} - e^{-(u-b-c)/T_2}) \qquad (u > b + c) \tag{A10.1.15}$$

对于脉冲式输入，在离散时刻 $t, t-1, t-2, \cdots$, 的输出，将通过适当选取参数值的差分方程

$$(1 + \xi_1\nabla + \xi_2\nabla^2)Y_t = g(1 + \eta_1\nabla + \eta_2\nabla^2)X_{t-b-1+} \tag{A10.1.16}$$

与输入联系起来. 上述差分方程也可写成

$$(1 - \delta_1 B - \delta_2 B^2)Y_t = (\omega_0 - \omega_1 B - \omega_2 B^2)X_{t-b-1+}$$

或

$$(1-S_1B)(1-S_2B)Y_t=(\omega_0-\omega_1B-\omega_2B^2)X_{t-b-1+} \quad \text{(A10.1.17)}$$

利用 (A10.1.1) 和 (A10.1.2), 并记.

$$S_1=e^{-1/T_1}\ , \qquad S_2=e^{-1/T_2}$$

我们得到

$$v_b=\int_{b+c}^{b+1}v(u)du=g(T_1-T_2)^{-1}[T_1(1-S_1^{1-c})-T_2(1-S_2^{1-c})]$$

$$v_{b+j}=\int_{b+j}^{b+j+1}v(u)du=g(T_1-T_2)^{-1}[T_1S_1^{-c}(1-S_1)S_1^j$$
$$-T_2S_2^{-c}(1-S_2)S_2^j] \qquad (j>1)$$

因此

$$(T_1-T_2)v(B)=gB^bT_1[1-S_1^{1-c}+S_1^{-c}(1-S_1)(1-S_1B)^{-1}S_1B]$$
$$-gB^bT_2[1-S_2^{1-c}+S_2^{-c}(1-S_2)(1-S_2B)^{-1}S_2B]$$

然而，由 (A10.1.17) 有

$$v(B)=\frac{B^b(\omega_0-\omega_1B-\omega_2B^2)}{(1-S_1B)(1-S_2B)}$$

我们得到

$$\begin{aligned}\omega_0&=g(T_1-T_2)^{-1}[T_1(1-S_1^{1-c})-T_2(1-S_2^{1-c})]\\ \omega_1&=g(T_1-T_2)^{-1}[(S_1+S_2)(T_1-T_2)\\ &\quad+T_2S_2^{1-c}(1+S_1)-T_1S_1^{1-c}(1+S_2)]\\ \omega_2&=gS_1S_2(T_1-T_2)^{-1}[T_2(1-S_1^{-c})-T_1(1-S_2^{-c}]\end{aligned} \quad \text{(A10.1.18)}$$

及

$$\begin{aligned}\delta_1&=S_1+S_2=e^{-1/T_1}+e^{-1/T_2}\\ \delta_2&=-S_1S_2=-e^{-(1/T_1)-(1/T_2)}\end{aligned} \quad \text{(A10.1.19)}$$

复根　如果 T_1 和 T_2 为复值，将

$$T_1 = \zeta^{-1}e^{i\lambda} \qquad T_2 = \zeta^{-1}e^{-i\lambda} \qquad (i^2 = -1)$$

代入可得到相应的表达式

$$\begin{aligned}
\omega_0 &= g\left\{1 - \frac{e^{-\zeta(1-c)\cos\lambda}\sin[\zeta(1-c)\sin\lambda + \lambda]}{\sin\lambda}\right\} \\
\omega_2 &= g\delta_2\left[1 - \frac{e^{\zeta c\cos\lambda}\sin(-\zeta c\sin\lambda + \lambda)}{\sin\lambda}\right] \\
\omega_1 &= \omega_0 - \omega_2 - (1 - \delta_1 - \delta_2)g
\end{aligned} \tag{A10.1.20}$$

这里

$$\begin{aligned}
\delta_1 &= 2e^{-\zeta\cos\lambda}\cos(\zeta\sin\lambda) \\
\delta_2 &= -e^{-2\zeta\cos\lambda}
\end{aligned} \tag{A10.1.21}$$

A10.2　非线性传递函数与线性化

我们前面所考虑传递函数的线性性质 (或可加性) 意味着：对若干个别输入之和的总响应就是对这些输入的个别响应之和。具体地，如果 $Y_t^{(1)}$ 是在时刻 t 对历史输入 $\{X_t^{(1)}\}$ 的响应，$Y_t^{(2)}$ 是在时刻 t 对历史输入 $\{X_t^{(2)}\}$ 的响应，那么，在时刻 t 对历史输入 $\{X_t^{(1)} + X_t^{(2)}\}$ 的响应就是 $Y_t^{(1)} + Y_t^{(2)}$。对于连续的输入和输出情况也完全类似。特别地，如果输入水平乘以某个常数，则输出水平也乘以同一常数。在实际中，这种假设不可能完全正确，但它对许多实际情形给出了有用的近似。

如果允许参数按某种限定方式依赖于输入的水平，有时便可得到非线性系统的模型。例如，假设一个系统在被研究的范围内 Y 有最大值 η，且对于任意 X 的稳态关系可以由二次表达式

$$Y_\infty = \eta - \frac{1}{2}k(\mu - X)^2$$

来近似。如前所述，这里的 Y 和 X 是从适当原点的偏离。那么

$$g(X) = \frac{dY_\infty}{dX} = k(\mu - X)$$

系统的动态特性能够用一阶差分方程 (10.3.4) 来描述，但是其中的增益是可变的，它与 $k(\mu - X)$ 成比例。于是便有

$$Y_t = \delta Y_{t-1} + k(\mu - X_{t-1+})(1-\delta)X_{t-1+} \quad \text{(A10.2.1)}$$

一个简单化学反应器的动态特性　有时会遇到这样的情形：我们对一个物理问题进行理论分析，就可以得出传递函数的恰当形式。特别是这样可以使我们能非常具体地看到，在非线性化近似中涉及到什么问题。

作为一个例子，假设将某种纯净化学物质 A 连续地馈入一个搅拌箱反应器中，当催化剂存在时，A 的某一比例变成了产品 B，但总的体积是不改变的；因此，连续地从反应器中离开的物质是由 B 和未变化的 A 混合组成。

假设系统最初是均衡的，其数量用适当的单位加以测量：

1. μ 是 A 馈入反应器的速度 (因而也是 A 和 B 的混合物离开反应器的速度)。

2. η 是在出口处未变化的 A 所占的比例，从而 $1-\eta$ 是在出口处产品 B 所占的比例。

3. V 是反应器的体积。

4. k 是一个常数，它决定了产品 B 的生成速度。

假设反应器关于 A 是“一阶”的，这是指 B 的生成速度和 A 的消耗速度都与 A 的当前总量成比例。那么 B 生成的速度为 $kV\eta$，而 B 离开出口的速度是 $\mu(1-\eta)$，由于系统是均衡的，所以

$$\mu(1-\eta) = kV\eta \quad \text{(A10.2.2)}$$

现在假设系统的均衡受到了干扰，在时刻 t 馈入反应器的速度是 $\mu + X(t)$，而在出口 A 的相应浓度是 $\eta + Y(t)$。现在 B 的化

学生成速度等于 $kV(\eta+Y(t))$，而 B 流出系统的速度现在等于 $[\mu+X(t)][1-\eta-Y(t)]$，一般这两个速度不再平衡了. 这两个量之差就是反应器中 B 的总量的增长速度，它等于 $-V[dY(t)/dt]$. 因此

$$-V\frac{dY(t)}{dt}=kV[\eta+Y(t)]-[\mu+X(t)][1-\eta-Y(t)] \quad \text{(A10.2.3)}$$

利用 (A10.2.2), 经重新整理， (A10.2.3) 可以写成

$$(kV+\mu+VD)Y(t)=X(t)[1-\eta-Y(t)]$$

或

$$(1+TD)Y(t)=g\left(1-\frac{Y(t)}{1-\eta}\right)X(t) \quad \text{(A10.2.4)}$$

其中

$$T=\frac{V}{kV+\mu} \quad , \quad g=\frac{1-\eta}{kV+\mu} \quad \text{(A10.2.5)}$$

于是 (A10.2.4) 是一个非线性微分方程，这是因为它包含 $X(t)$ 乘以 $Y(t)$ 这项. 但在有些实际情况中，它可以由一个线性微分方程恰当地近似，我们下面就加以说明.

过程的运作可以在范围很广的条件下进行，但是如下情形无疑不能算是异乎寻常的: 馈入量 A 转化为产品 B 的百分比 $100(1-\eta)$ 假设为 80%, 具有实际意义的波动百分比 $100Y(t)$ 为 4% . 在这种情形，系数 $1-Y(t)/(1-\eta)$ 在 0.95 到 1.00 之间变化，十分接近于 1, 因而可用 1 来代替. 由此，非线性微分方程 (A10.2.4) 可用线性一阶微分方程

$$(1+TD)Y(t)=gX(t)$$

来代替，其中 T 和 g 的定义如 10.1.2 节. 如果系统是在离散的时间间隔上观测的，该方程还可由一个线性差方程来近似.

非线性成为重要问题的情形显然是会发生的，在最优化的研究中，当变量可以在很大的范围变化时，尤其是这样．在线性假设不恰当时，有一种方法是很有用的，那就是用适合于输入变量不同区间的一组线性模型来描述动态特性．但是对于离散系统而言，假若必须直接处理非线性差分方程，则递推“求解”比解析方法更方便些．例如，对于非线性微分方程 (A10.2.4)，我们可以代之以非线性差分方程

$$(1+\xi_1\nabla)Y_t = g(1+\eta_{12}Y_{t-1})X_{t-1}$$

11

传递函数模型的识别、拟合及检验

在第 10 章中我们引入了一组参数简约的线性传递函数模型

$$Y_t - \delta_1 Y_{t-1} - \cdots - \delta_r Y_{t-r} = \omega_0 X_{t-b} - \omega_1 X_{t-b-1} - \cdots - \omega_s X_{t-b-s}$$

或

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b}$$

在这些模型中 X_t 和 Y_t 是系统的输入和输出对均衡的偏离。在实际中，系统会受到扰动或噪声的影响，其直接后果是在传递函数模型所预报的输出中搀入了一个量 N_t 。传递函数－噪声联合模型可以写为

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t$$

本章中所讲述的是：当在离散等间隔时刻 $1, 2, \cdots, N$ 上，得到了输入和输出的同时观测对 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_N, Y_N)$, 在此基础上，我们对于传递函数－噪声模型的识别、拟合及检验方法。

工程上对于传递函数的估计通常是基于对系统特定输入的选取，比如阶跃输入、正弦波输入 [211] 和“脉冲”式输入 [114] 。这些方法在系统只受少量噪声影响时是有效的，但在其它情形就不能令人满意。当存在不可忽视的噪声时，必须使用统计方法来估计传递函数。以往曾有两种方法尝试解决这一问题，一是在时域中直接估计脉冲响应；另一是在频域直接估计增益和相位特征，

如见 [62],[115] 和 [122] 。这些方法常常不能令人满意，因为它们都涉及大量参数的估计。例如，为了确定增益和相位特征必须要在每个频率上估计两个参数。本章所采用的方法是估计简约差分方程模型中的参数。在贯穿本章的绝大部分内容中，我们假设输入 X_t 本身是一个随机过程。这种模型在描述和预报某些多维时间序列时是有用的。

11.1 互相关函数

和自相关函数用于识别随机模型一样，在传递函数模型的识别中使用的分析工具是输入和输出之间的互相关函数。在本节中我们描述互相关函数的基本特征，下一节说明如何将它用于识别传递函数模型。

11.1.1 互协方差和互相关函数的性质

双变量随机过程　在第 2 章中我们已经看到，为了分析一个统计时间序列，我们将它看作是一个假想时间序列总体 (称之为随机过程) 的实现，这样做是有益的。现在，假设我们希望描述某个物理系统的输入时间序列，及相应的输出时间序列。例如，图 11.1 给出的连续数据表示输入煤气的馈入速度 (数码化的) 和相应地从煤气炉输出 CO_2 的浓度。随后，我们可以把这一对时间序列看作是一对设定时间序列总体 —双变量随机过程 (X_t, Y_t) 的实现。我们假设这些数据是在等间隔时刻读到的，从而得到由双变量过程产生的一对离散时间序列，时间序列在时刻 $t_0+h, t_0+2h, \cdots, t_0+Nh$ 的值记作 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_N, Y_N)$ 。

在本章中，大部分说明使用的是以 9 秒为间隔读到的煤气炉数据 (见图 11.1) 。由此而获得的值 (X_t, Y_t) 列在第五部分“时间序列汇集”中，叫做序列 J 。

互协方差和互相关函数　在第 2 章中我们已经看到，一个

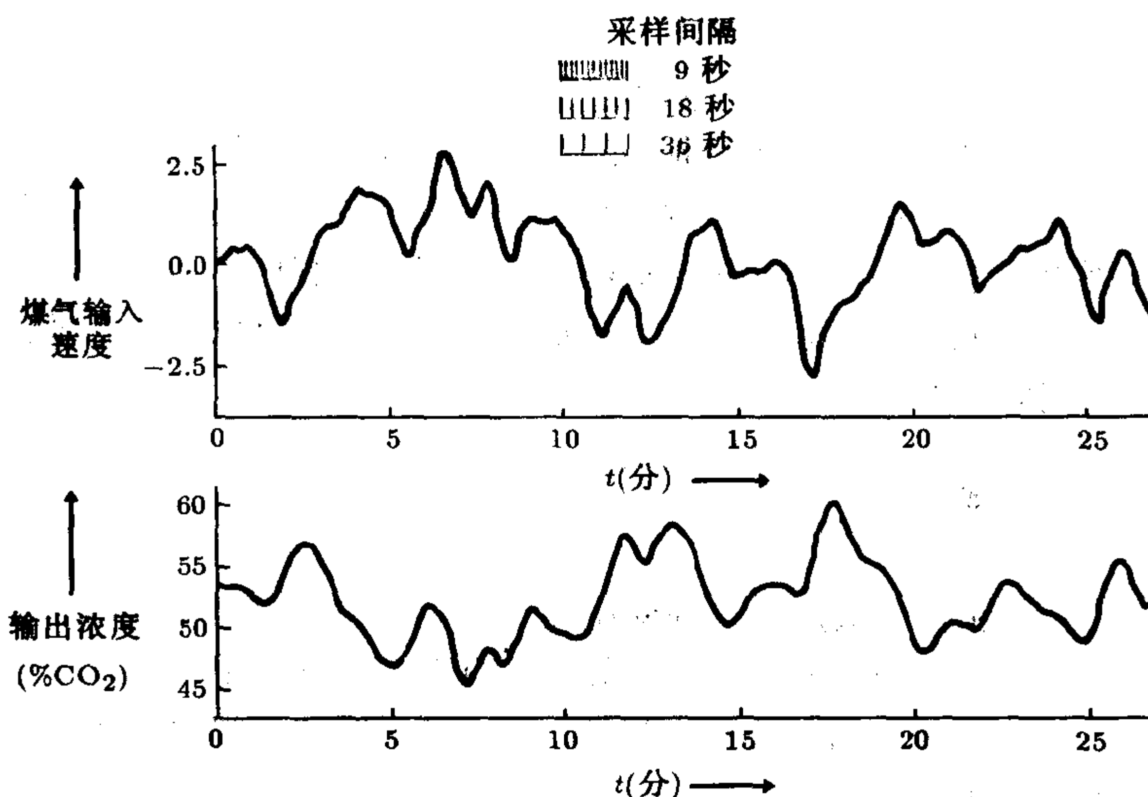


图 11.1 气炉的煤气输入速度和 CO_2 的输出浓度

平稳高斯随机过程可以用其均值 μ 和自协方差函数 γ_k 来描述，或者等价地，用其均值 μ，方差 σ^2 和自相关函数 ρ_k 来描述。而且，由于 $\gamma_k=\gamma_{-k}$ 及 $\rho_k=\rho_{-k}$，因此自协方差和自相关函数只需对非负的滞后值 $k=0,1,2\cdots$ 画出。

一般地，一个双变量随机过程无须是平稳的，但是，和第 4 章中一样，我们假设适当差分的过程 (x_t,y_t) 是平稳的，这里 $x_t=\nabla^d X_t, y_t=\nabla^d Y_t$。这一平稳假设特别意味着分量过程 x_t 和 y_t 有不变的均值 μ_x 和 μ_y，及不变的方差 σ_x^2 和 σ_y^2。如果另外还假定该双变量过程是高斯 (或正态) 的，那么其特征被均值 μ_x,μ_y 和协方差阵唯一地描述。图 11.2 给出了需要考虑的各种不同的协方差。

每个分量序列滞后 k 的自协方差系数一般由下面公式定义:

$$\gamma_{xx}(k)=E[(x_t-\mu_x)(x_{t+k}-\mu_x)]=E[(x_t-\mu_x)(x_{t-k}-\mu_x)]$$
$$\gamma_{yy}(k)=E[(y_t-\mu_y)(y_{t+k}-\mu_y)]=E[(y_t-\mu_y)(y_{t-k}-\mu_y)]$$

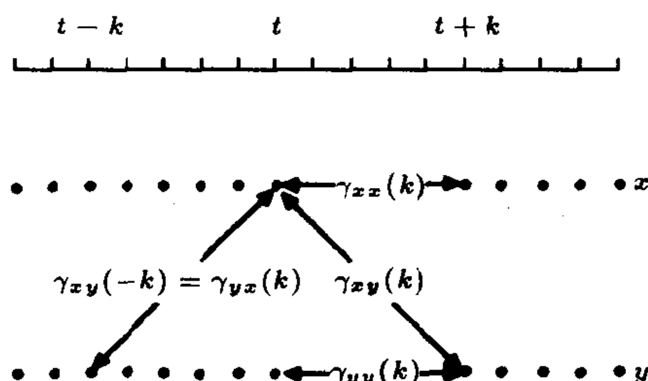


图 11.2　双变量随机过程的自协方差和互协方差

这里我们用推广的符号 $\gamma_{xx}(k)$ 和 $\gamma_{yy}(k)$ 来记 x 和 y 的自协方差。在协方差阵中可能出现的其它协方差只有 x 和 y 之间滞后为 $+k$ 的互协方差:

$$\gamma_{xy}(k) = E[(x_t - \mu_x)(y_{t+k} - \mu_y)] \qquad (k = 0, 1, 2, \cdots) \qquad (11.1.1)$$

y 和 x 之间滞后为 $+k$ 的互协方差:

$$\gamma_{yx}(k) = E[(y_t - \mu_y)(x_{t+k} - \mu_x)] \qquad (k = 0, 1, 2, \cdots) \qquad (11.1.2)$$

注意, $\gamma_{xy}(k)$ 和 $\gamma_{yx}(k)$ 一般并不相同，但是，由于

$$\gamma_{xy}(k) = E[(x_{t-k}-\mu_x)(y_t-\mu_y)] = E[(y_t-\mu_y)(x_{t-k}-\mu_x)] = \gamma_{yx}(-k)$$

我们只需要对 $k = 0, \pm1, \pm2, \cdots$ 定义函数 $\gamma_{xy}(k)$, 并称之为双变量过程的互协方差函数。类似地，无量纲的量

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \qquad k = 0, \pm1, \pm2, \cdots \qquad (11.1.3)$$

称为滞后 k 的互相关系数，而函数 $\rho_{xy}(k), k = 0, \pm1, \pm2, \cdots$ 称为双变量过程互相关函数。

由于 $\rho_{xy}(k)$ 一般不等于 $\rho_{xy}(-k)$, 互相关函数与自相关函数不同，它关于 $k = 0$ 不是对称的。事实上经常遇到的情形是：互相关函数在 $-\infty$ 到 i 或 i 到 $+\infty$ 的整个区间上都为零。例如，对有延迟的一阶自回归过程

$$(1 - \phi B)\widetilde{z}_t = a_{t-b} \qquad (-1 < \phi < 1,\ b > 0)$$

考虑 a 和 z 之间的自协方差函数，这里 a_t 是零均值的。由于

$$\widetilde{z}_{t+k} = a_{t+k-b} + \phi a_{t+k-b-1} + \phi^2 a_{t+k-b-2} + \cdots$$

因此 a 和 z_t 之间的协方差函数是

$$\gamma_{az}(k) = E[a_t \widetilde{z}_{t+k}] = \begin{cases} \phi^{k-b}\sigma_a^2 & ,k \geqslant b \\ 0 & ,k < b \end{cases}$$

从而对于有延迟自回归过程，其互相关函数为

$$\rho_{az}(k) = \begin{cases} \phi^{k-b}\frac{\sigma_a}{\sigma_z} = \phi^{k-b}(1-\phi^2)^{1/2} & , \quad k \geqslant b \\ 0 & , \quad k < b \end{cases}$$

图 11.3 给出了 $b=2$ 和 $\phi=0.6$ 的互相关函数。

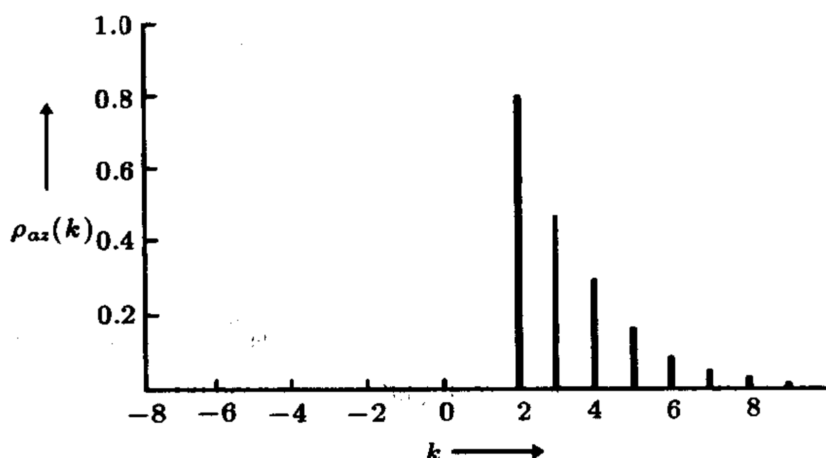


图 11.3 有延迟过程 $\widetilde{z}_t - 0.6\widetilde{z}_{t-1} = a_{t-2}$ 的 a 和 z 之间的互相关函数

11.1.2 互协方差和互相关函数的估计

假设对原始的输入和输出时间序列作 d 次差分后有 $n = N-d$ 对值 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ 可供分析。那么可以证明 (例如在 [122] 中)，滞后 k 互协方差系数的估计值 $c_{xy}(k)$ 由下式给出

$$c_{xy}(k) = \begin{cases} \dfrac{1}{n}\sum\limits_{t=1}^{n-k}(x_t - \overline{x})(y_{t+k} - \overline{y}) \;, & k = 0,1,2,\cdots \\ \dfrac{1}{n}\sum\limits_{t=1}^{n+k}(y_t - \overline{y})(x_{t-k} - \overline{x}) \;, & k = 0,-1,-2,\cdots \end{cases} \tag{11.1.4}$$

这里 $\overline{x}$ 和 $\overline{y}$ 分别是序列 x 和序列 y 的样本均值。类似地，滞后 k 互相关系数 $\rho_{xy}(k)$ 的估计值 $r_{xy}(k)$ 可以得到，只须在 (11.13) 中用估计值 $c_{xy}(k)$ 代替 $\gamma_{xy}(k)$，用 $s_x = \sqrt{c_{xx}(0)}$ 代替 σ_x 用 $s_y = \sqrt{c_{yy}(0)}$ 代替 σ_y，于是得出

$$r_{xy}(k) = \frac{c_{xy}(k)}{s_x s_y} \qquad (k = 0, \pm 1, \pm 2, \cdots) \tag{11.1.5}$$

例： 在实际中为了获得互相关函数有意义的估计，我们至少需要 50 对观测值。但是为了对公式 (11.1.4) 和 (11.1.5) 加以说明，在计算滞后为 1 和 -1 的互相关函数估计值时，只对下面 5 对观测值的序列进行计算

t	1	2	3	4	5
x_t	11	7	8	12	14
y_t	7	10	6	7	10

现在 $\overline{x} = 10.4, \overline{y} = 8$，故对均值的离差为

t	1	2	3	4	5
$x_t - \overline{x}$	-0.6	-3.4	-2.4	1.6	3.6
$y_t - \overline{y}$	-1.0	2.0	-2.0	-1.0	2.0

因此

$$\begin{aligned}
&\sum_{t=1}^{4}(x_t - \overline{x})(y_{t+1} - \overline{y}) \\
&= (0.6)(2.0) + (-3.4)(-1.0) + (-2.4)(-1.0) + (1.6)(2.0) \\
&= 13.60
\end{aligned}$$

并且 $c_{xy}(1) = 13.60/5 = 2.720$ 。利用 $s_x = 2.577$ 和 $s_y = 1.673$，我们得到

$$r_{xy}(1) = \frac{c_{xy}(1)}{s_x s_y} = \frac{2.720}{(2.577)(1.673)} = 0.63$$

类似地，$\sum_{t=1}^{4}(y_t - \overline{y})(x_{t+1} - \overline{x}) = -8.20$ 。从而 $c_{xy}(-1) = -1.64$，且

$$r_{xy}(-1) = \frac{-1.640}{(2.577)(1.673} = -0.38$$

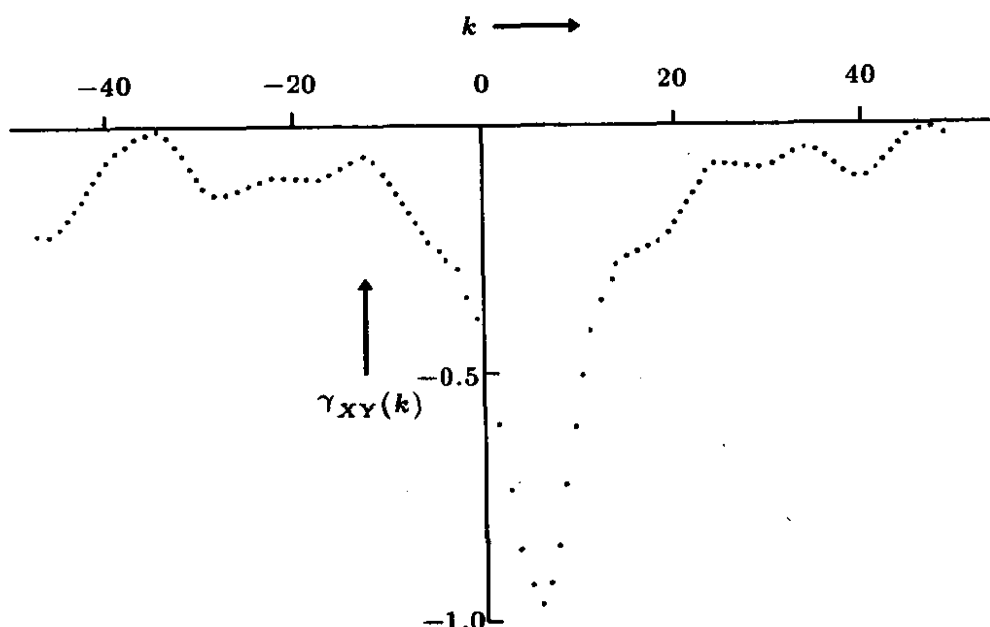


图 11.4　以 9 秒为间隔读取代码化煤气炉数据，其输入和输出之间的互相关函数

图 11.4 给出了估计出来的离散煤气炉数据输入和输出之间的互相关函数 $r_{xy}(k)$，这些数据是从图 11.1 的连续资料中每隔 9 秒读出的。我们看到，互相关函数关于零点不是对称的，在 $k=+5$ 有明显的峰，这表明输出滞后于输入。且这个互相关为负值，这是可以预料的，因为 (见图 11.1) 数码化输入的增加会导致输出值的减少。

11.1.3　互相关估计的近似标准差

为了检验互相关函数 $\rho_{xy}(k)$ 的某些值为零是否有效，一个粗略的方法是：把相应的互相关估计值与从 Bartlett 公式 [26] 得出的它们的近似标准差作比较。Bartlett 证明，在正态假设下，两个互相关估计 $r_{xy}(k)$ 和 $r_{xy}(k+l)$ 之间的方差，当 $k>0$ 时由下式给出：

$$\operatorname{cov}[r_{xy}(k),r_{xy}(k+l)]$$
$$\simeq (n-k)^{-1}\sum_{v=-\infty}^{+\infty}\{\rho_{xx}(v)\rho_{yy}(v+l)+\rho_{xy}(-v)\rho_{xy}(v+2k+l)$$

$$
\begin{aligned}
&+\rho_{xy}(k)\rho_{xy}(k+l)[\rho_{xy}^2(v)+\frac{1}{2}\rho_{xx}^2(v)+\frac{1}{2}\rho_{yy}^2(v)]\\
&-\rho_{xy}(k)[\rho_{xx}(v)\rho_{xy}(v+k+l)+\rho_{xy}(-v)\rho_{yy}(v+k+l)]\\
&-\rho_{xy}(k+l)[\rho_{xx}(v)\rho_{xy}(v+k)+\rho_{xy}(-v)\rho_{yy}(v+k)]\}
\end{aligned} \tag{11.1.6}
$$

特别，令 $l=0$, 有

$$
\begin{aligned}
\operatorname{var}[r_{xy}(k)] \simeq (n-k)^{-1} \sum_{v=-\infty}^{+\infty} \{&\rho_{xx}(v)\rho_{yy}(v)+\rho_{xy}(k+v)\rho_{xy}(k-v)\\
&+\rho_{xy}^2[\rho_{xy}^2(v)+\frac{1}{2}\rho_{xx}^2(v)+\frac{1}{2}\rho_{yy}(v)]\\
&-2\rho_{xy}(k)[\rho_{xx}(v)\rho_{xy}(v+k)+\rho_{xy}(-v)\rho_{yy}(v+k)]\}
\end{aligned} \tag{11.1.7}
$$

正如 Bartlett 所指出，由这些一般表达式可以推导出应用于重要特殊情形的公式. 例如若假设 $x_t \equiv y_t$, 只需令

$$
\rho_{xx}(v)=\rho_{yy}(v)=\rho_{xy}(v)=\rho_{xy}(-v)
$$

即可. 将该式代入 (11.1.6) 和 (11.1.7) 就得到两个自相关估计之间协方差的表达式，更特殊的情形是自相关估计方差的表达式，这已由 (2.1.11) 给出.

常常会有这样的情形：两个过程仅在相当窄的滞后范围内显著地互相关。若 $\rho_{xy}(v)$ 仅在某个范围 $Q_1 < b < Q_2$ 非零，那么

1. 如果 $k,k+l$ 和 $k+\frac{1}{2}l$ 都不在这一范围内，(11.1.6) 中除了第一项以外全都为零，且

$$
\operatorname{cov}[r_{xy}(k), r_{xy}(k+l)] \simeq (n-k)^{-1} \sum_{-\infty}^{\infty} \rho_{xx}(v)\rho_{yy}(v+l) \tag{11.1.8}
$$

2. 如果 k 不在该范围内，用类似方法，(11.1.7) 化简为

$$
\operatorname{var}[r_{xy}(k)] \simeq (n-k)^{-1} \sum_{v=-\infty}^{\infty} \rho_{xx}(v)\rho_{yy}(v) \tag{11.1.9}
$$

特别地，在两过程无互相关这一假设之下，简化公式适用于所有滞后 k 和 $k+l$ 。另一种较有意义的特殊情形是：两过程不存在

互相关关系，且其中一个是白噪声．假设 $y_t = a_t$ 是由白噪声过程生成的，而 x_t 是自相关的，那么由 (11.1.8),

$$\text{cov}[r_{xa}(k), r_{xa}(k+l)] \simeq (n-k)^{-1}\rho_{xx}(l) \tag{11.1.10}$$

$$\text{var}[r_{xa}(k)] \simeq (n-k)^{-1} \tag{11.1.11}$$

由此得到

$$\rho[r_{xa}(k), r_{xa}(k+l)] \simeq \rho_{xx}(l) \tag{11.1.12}$$

因此在这种情形，两过程的互相关与生成 x_t 的过程具有相同的自相关．所以，尽管 a_t 和 x_t 是不相关的，可以预料其互相关函数在零附近变动，标准差为 $(n-k)^{-1/2}$，且具有自相关函数 $\rho_{xx}(l)$ 典型特征的规律性状．最后，如果两个过程都是白噪声且互不相关，则互相关之间的协方差将为零．

11.2 传递函数模型的识别

我们现在说明如何识别传递函数 - 噪声混合模型

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t$$

这是一个输出中混有噪声 N_t 的线性系统，假定噪声是由 ARIMA 过程产生且与输入 X_t 独立[①]，具体说来，这一阶段的目的是得出有关传递函数模型算子的分母和分子阶数 r 和 s 的一些设想，并由此引伸得到关于参数 $\boldsymbol{\delta}, \boldsymbol{\omega}$ 和延迟参数 b 的初步猜测．另外，我们的目的是：对于描述输出中噪声的 ARIMA 模型参数 p, d, q 做出粗略的猜测，并得出该模型中参数 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的初估计．由此得到试探性的传递函数 - 噪声模型可以作为节 11.3 中所述更有效估计方法的出发点．

[①] 当输入可由我们选择时，我们可以确保 X_t 按照某随机过程生成，且与 N_t 独立．

识别方法概要　假设传递函数模型

$$Y_t = v(B)X_t + N_t \tag{11.2.1}$$

可有如下的简约参数化形式

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t \tag{11.2.2}$$

这里 $\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r$ 及 $\omega(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$ 。识别程序如下：

1. 得出 (11.2.1) 中脉冲响应权数 v_j 的初估计 $\hat{v}_j$ 。

2. 利用得到的估计 $\hat{v}_j$ 来猜测 (11.2.2) 中分母和分子的算子阶数 r 和 s, 以及延迟参数 b 。

3. 将估计值 $\hat{v}_j$ 代入等式 (10.2.8),r, s 和 b 的值由 (2) 中获得. 这样可得到 (11.2.2) 中参数 $\boldsymbol{\delta}$ 和 $\boldsymbol{\omega}$ 的初估计.

有了 $\hat{v}_j, b, r$ 和 s 的值，就可用节 10.2.2 中确立的方法加以猜测，即形如 (11.2.2) 的模型，其脉冲响应权数的组成应为：

1. b 个零值 $v_0, b_1, \cdots, v_{b-1}$ 。

2. 随后的 $s - r + 1$ 个值 $v_b, v_{b+1}, \cdots, v_{b+s-r}$ 没有固定的形式 (若 $s < r$ 则不出现这样的值) 。

3. $j > b + s - r + 1$ 时的 v_j 值遵从 r 阶差分方程， r 个初值为 $v_{b+s}, \cdots, v_{b+s-r+1}$ 。当 $j < b$ 时的初值 v_j 当然为零。

对输入和输出的差分　在过程识别中使用的基本工具是输入和输出之间的互相关函数。当过程为非平稳时，则假设经适当差分后可以达到平稳。如果由序列 (X_t, Y_t) 估计出的自相关和互相关函数不是很快地衰减，就很有可能具有非平稳特性。我们假定平稳化所需的 d 阶差分 [①] 实施之后，估计出的 $x_t = \nabla^d X_t$ 和 $y_t = \nabla^d Y_t$ 的自相关 $r_{xx}(k)$ 和， $r_{yy}(k)$ 和互相关 $r_{xy}(k)$ 都很快地衰减。实际中 d 通常为 0,1 或 2 。

---

① 这里所述方法对于输入和输出使用不同差分阶数时仍然有效。

非预白噪化的脉冲响应函数的识别　假设 d 次差分后，模型 (11.2.1) 可写为形如

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \cdots + n_t \tag{11.2.3}$$

这里 $y_t = \nabla^d Y_t, x_t = \nabla^d X_t$ 和 $n_t = \nabla^d N_t$ 是零均值平稳过程. 那么，(11.2.3) 两边乘以 $x_{t-k}, k > 0$, 我们得到

$$x_{t-k} y_t = v_0 x_{t-k} x_t + v_1 x_{t-k} x_{t-1} + \cdots + x_{t-k} n_t \tag{11.2.4}$$

如果我们进一步假设对于所有的 k 有 x_{t-k} 和 n_t 不相关，对 (11.2.4) 取期望可得出方程组

$$\gamma_{xy}(k) = v_0 \gamma_{xx}(k) + v_1 \gamma_{xx}(k-1) + \cdots \qquad (k = 0, 1, 2, \cdots) \tag{11.2.5}$$

假定在 $k = K$ 之后，权数 v_j 等效于零，那么方程 (11.2.5) 中的前 $K+1$ 个可写成

$$\boldsymbol{\gamma}_{xy} = \boldsymbol{\Gamma}_{xx} \boldsymbol{v} \tag{11.2.6}$$

其中

$$\boldsymbol{\gamma}_{xy} = \begin{bmatrix} \gamma_{xy}(0) \\ \gamma_{xy}(1) \\ \vdots \\ \gamma_{xy}(K) \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_K \end{bmatrix}$$

$$\boldsymbol{\Gamma}_{xx} = \begin{bmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \cdots & \gamma_{xx}(K) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \cdots & \gamma_{xx}(K-1) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \gamma_{xx}(K) & \gamma_{xx}(K-1) & \cdots & \gamma_{xx}(0) \end{bmatrix}$$

把输入的自相关函数估计值 $c_{xx}(k)$ 和输入、输出之间的互相关估计值 $c_{xy}(k)$ 代入，(11.2.6) 给出了前 $K+1$ 个权数的 $K+1$ 个线性方程。但这些方程一般不能提供有效的估计，因为求解十分麻

烦，而且在任何情形都必须知道 v_j 在其后等效于零的 K 点。方程组 (11.2.6) 的样本形式实际上表示了 y_t 对 $x_t, x_{t-1}, \cdots, x_{t-k}$ 做线性回归的最小二乘正规方程组，且隐含假定 (11.2.3) 中的噪声 n_t 不是自相关的。这正是这种回归识别方法无效的一个原因。为了改进这种方法的效果，Liu 和 Hanssens[217](也可见 Pankratz[218,Ch. 5]) 建议对回归方程 $y_t = v_0x_1 + v_1x_{t-1} + \cdots + v_kx_{t-k} + n_t$ 进行广义最小二乘估计，并假设噪声 n_t 遵从某个相关的时间序列 ARMA 模型。他们还讨论了这种脉冲响应函数识别方法的推广，将其用于模型中有多个输入过程 $X_{1,t}, X_{2,t}, \cdots, X_{m,t}$ 的情形，即 $Y_t = v_1(B)X_{1,t} + \cdots + v_m(B)X_{m,t} + N_t$ 。

11.2.1 对预白噪化输入传递函数模型的识别

如果系统的输入是白噪声，则识别过程将很大程度被简化。正如节 11.6 中更详细地讨论，当输入的选择可由我们安排，采取什么样的输入就要多加斟酌了。当原始输入遵从某种别随机过程，通过白噪化就可使问题简化。

假定适当差分之后的输入过程是平稳的，并且可由自回归滑动平均的广义线性类中的某些模型来描述。那么，给定一组数据，我们可以对 x_t 用通常的识别和估计方法得到一个模型：

$$\theta_x^{-1}(B)\phi_x(B)x_t = \alpha_t \tag{11.2.7}$$

该式把相关的输入序列 x_t 转化为非常接近不相关的白噪声 α_t 。同时我们可以由 $\hat{\alpha}$ 的平方和得到 σ_α^2 的估计值 s_α^2 。如果我们现在对 y_t 使用同样的变换得到

$$\beta_t = \theta_x^{-1}(B)\phi_x(B)y_t$$

该模型可以写为

$$\beta_t = v(B)\alpha_t + \epsilon_t \tag{11.2.8}$$

这里， ϵ_t 是转换后的噪声序列，由下式定义

$$\epsilon_t = \theta_x^{-1}(B)\phi_x(B)n_t \tag{11.2.9}$$

在 (11.2.8) 两边同乘以 α_{t-k} 并取期望，我们得到

$$\gamma_{\alpha\beta}(k) = v_k \sigma_\alpha^2 \tag{11.2.10}$$

这里，$\gamma_{\alpha\beta}(k) = E[\alpha_{t-k}\beta_t]$ 是 α 和 β 之间滞后为 $+k$ 的互协方差。于是

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2}$$

或者用互相关来表示

$$v_k = \frac{\rho_{\alpha\beta}(k)\sigma_\beta}{\sigma_\alpha} \qquad (k = 0, 1, 2, \cdots) \tag{11.2.11}$$

因此，将输入白噪化后，在白噪化输入和经相应变换的输出之间的互相关函数直接与脉冲响应函数成正比。我们注意到，白噪化的效果是将非正交的方程组 (11.2.6) 转化为正交方程组 (11.2.10) 。

在实际中，我们并不知道理论互相关函数 $\rho_{\alpha\beta}(k)$，故我们必须将其代入 (11.2.11)，得到

$$\hat{v}_k = \frac{r_{\alpha\beta}(k)s_\beta}{s_\alpha} \qquad (k = 0, 1, 2, \cdots) \tag{11.2.12}$$

由此获得的初步估计 $\hat{v}_k$ 一般仍是统计上无效的，但是可以为在传递函数模型中选择适当的算子 $\delta(B)$ 和 $\omega(B)$ 提供大略的基础。现在我们用一个实际例子来说明这种识别和初步估计的方法。

11.2.2 传递函数模型识别的例子

在自适应的优化研究中 [136]，涉及到一个煤气炉，炉中是空气和甲烷混合而成的含 CO_2 (二氧化碳) 的煤气混合物。空气的馈入保持恒定，但甲烷的馈入速度可根据任何需要的方式改变，在输出煤气中生成 CO_2 的浓度是可以测量的。图 11.1 中连续数据的采集给出了我们感兴趣的范围内系统的动态特性，在此范围内，已知该系统适用于近似的线性稳态关系。在图 11.1 的上半

部，给出了连续的随机输入序列 $X(t)$，它是由过去的白噪声通过一个线性滤波器生成。这个过程的均值为零，并且在该实验的实现中其变动范围是 -2.5 到 2.5. 这时希望甲烷实际的馈入速度在每分钟 0.5 到 0.7 立方英尺的范围内。为确保这一点，输入煤气的馈入速度应服从如下过程

$$\text{甲烷馈入速度} = 0.60 - 0.04X(t)$$

为简单起见，我们将始终使用“数码化”的输入 $X(t)$ 。最终用实际输入速度表示的传递函数很容易由替换得到。第五部分“时间序列汇集”中的序列 J 给出了 296 个观测对 (X_t, Y_t) 的顺序值，它们是以 9 秒为间隔从连续记录中读出的。在这个具体例实验中，输入扰动的特征是已知的，因为我们曾详细地导出。但我们还是把它当作未知来处理。被估计的 X_t 和 Y_t 的自相关和互相关函数很快地衰减，这表明不需要再做差分。对 X_t 运用一般的检验和拟合表明可用一个 3 阶自回归过程进行很好的描述

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)X_t = \alpha_t$$

这里 $\hat{\phi}_1 = 1.97, \hat{\phi}_2 = -1.37, \hat{\phi}_3 = 0.34$, 及 $s_\alpha^2 = 0.0353$ 。因此，将变换

$$\alpha_t = (1 - 1.97B + 1.37B^2 - 0.34B^3)X_t$$
$$\beta_t = (1 - 1.97B + 1.37B^2 - 0.34B^3)Y_t$$

用于输入和输出序列，得出对于序列 α 和 β 有 $s_\alpha = 0.188, s_\beta = 0.358$ 。 α 和 β 之间互相关的估计在表 11.1 中给出，另外，还给出了脉冲响应函数的估计值 (11.2.12)

$$\hat{v}_k = \frac{0.358}{0.188} r_{\alpha\beta}(k)$$

表 11.1　煤气炉数据白噪化后的互相关函数和近似脉冲响应函数

k	$r_{\alpha\beta}(k)$	$\hat{\sigma}(r)$	$\hat{v}_k$	$r_{\beta\beta}(k)$	k	$r_{\alpha\beta}(k)$	$\hat{\sigma}(r)$	$\hat{v}_k$	$r_{\beta\beta}(k)$
0	−0.01	0.06	−0.02	1.00	6	−0.27	0.06	−0.52	0.12
1	0.05	0.06	0.10	0.22	7	−0.17	0.06	−0.32	0.04
2	−0.03	0.06	−0.06	0.36	8	−0.03	0.06	−0.06	0.09
3	−0.28	0.05	−0.53	0.13	9	0.03	0.06	0.06	0.00
4	−0.33	0.06	−0.63	0.08	10	−0.05	0.06	−0.10	0.10
5	−0.46	0.05	−0.88	0.01					

在表 11.1 中列出的互相关函数 $\gamma_{\alpha\beta}(k)$ 的近似标准差是从 (11.1.7) 得出的方差的平方根：

1. 滞后在 $+2$ 以下和 $+8$ 以上的互相关假设等于零。
2. $k>0$ 的自相关 $\rho_{\alpha\alpha}(k)$ 假设为零。
3. $k>4$ 的自相关 $\rho_{\beta\beta}(k)$ 假设为零。
4. 用表 11.1 中相关函数 $r_{\alpha\beta}(k)$ 和 $r_{\beta\beta}(k)$ 的估计值代替理论值。

表 11.1 还给出了 $r_{\beta\beta}(k)$ 的值，这是 (11.1.7) 所需要的。在图 11.5 中绘出了互相关的估计值以及以零为中心的一倍和两倍的标准差概率限。在这个例子中标准差与近似值 $n^{-1/2}=0.06$ 的差别很小，因此适用于序列不相关的假设。值 $\hat{v}_0,\hat{v}_1$ 和 $\hat{v}_2$ 与其标准差相比很小，这意味着 $b=3$(两个整的延迟期),$\hat{v}$ 的其它特性可以用一个模型来解释，其参数 (r,s,b) 等于 (1,2,3), 或者等于 (2,2,3)．第一个模型意味着 v_3 和 v_4 是没有固定形式的初值。差分方程 $v_j-\delta v_{j-1}=0, j>5$ 确定了指数衰减的变化， v_5 给出了起始值。第二个模型意味着 v_3 是唯一形式不定的初值，而 v_4 和 v_5 给出了差分方程 $v_j\delta_1 v_{j-1}-\delta_2 v_{j-2}=0, j>5$ 所确定的双重指数衰减的起始值。

这样，经初步识别给出了传递函数模型

$$(1-\delta_1 B-\delta_2 B^2)Y_t=(\omega_0-\omega_1 B-\omega_2 B^2)X_{t-b} \qquad (11.2.13)$$

或者是它的某种简化，如可能 $b=3$ 。

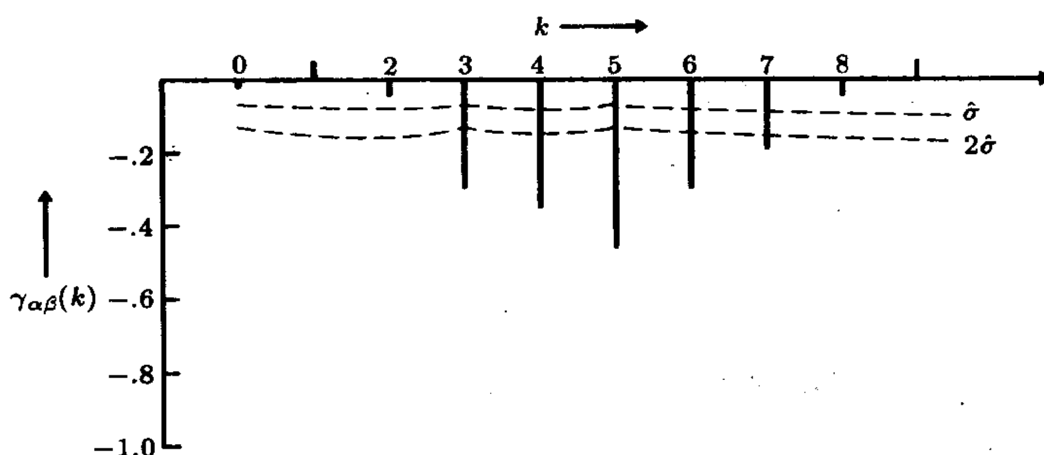


图 11.5　白噪化后煤气炉数据互相关函数的估计

初估计　假设在模型 (11.2.13) 中 $b=3$, 关于脉冲响应函数的方程 (10.2.8) 是

$$\begin{aligned}
v_j &= 0 \qquad (j<3) \\
v_3 &= \omega_0 \\
v_4 &= \delta_1 v_3 - \omega_1 \\
v_5 &= \delta_1 v_4 + \delta_2 v_3 - \omega_2 \\
v_6 &= \delta_1 v_5 + \delta_2 v_4 \\
v_7 &= \delta_1 v_6 + \delta_2 v_5
\end{aligned} \tag{11.2.14}$$

将表 11.1 中的估计值代入最后两个方程，我们得到

$$\begin{aligned}
-0.88\hat{\delta}_1 - 0.63\hat{\delta}_2 &= -0.52 \\
-0.52\hat{\delta}_1 - 0.88\hat{\delta}_2 &= -0.32
\end{aligned}$$

这给出了初估计 $\hat{\delta}_1=0.57, \hat{\delta}_2=0.02$ 。如果将这些值代入方程组 (11.2.14) 的第 2,3 和第 4 个方程，我们得到

$$\begin{aligned}
&\hat{\omega}_0 = \hat{v}_3 = -0.53 \\
&\hat{\omega}_1 = \hat{\delta}_1\hat{v}_3 - \hat{v}_4 = (0.57)(-0.53) + 0.63 = 0.33 \\
&\hat{\omega}_2 = \hat{\delta}_1\hat{v}_4 + \hat{\delta}_2\hat{v}_3 - \hat{v}_5 = (0.57)(-0.63) + (0.02)(-0.53) + 0.88 = 0.51
\end{aligned}$$

于是初步识别给出了一个尝试性的传递函数模型

$$(1-0.57B-0.02B^2)Y_t = -(0.53+0.33B+0.51B^2)X_{t-3}$$

由此获得的估值可以作为起始值以采用更有效的迭代方法，有关方法在节 11.3 中叙述。我们注意到， $\hat{\delta}_2$ 很小，这意味着该参数可以略去，但现在我们仍然保留它。

11.2.3 噪声模型的识别

回到一般情形，假设 (如果必要的话，做适当的差分之后) 模型可以写成

$$y_t = v(B)x_t + n_t$$

这里

$$n_t = \nabla^d N_t$$

用 11.2.2 节讨论的方法给出传递函数的初步估计之后，噪声序列的估计由下式得到

$$\hat{n}_t = y_t - \hat{v}(B)x_t$$

即

$$\hat{n}_t = y_t - \hat{v}_0 x_t - \hat{v}_1 x_{t-1} - \hat{v}_2 x_{t-2} - \cdots$$

另外， $\hat{v}(B)$ 还可以用初步识别所确定的试用传递函数模型来替代。于是

$$\hat{n}_t = y_t = \hat{\delta}^{-1}(B)\hat{\omega}(B)x_{t-b}$$

$\hat{n}_t$ 可以这样来计算：先通过把 $\hat{\delta}(B)\hat{y}_t = \hat{\omega}(B)x_{t-b}$ 写为

$$\hat{y}_t = \hat{\delta}_1\hat{y}_{t-1} + \cdots \hat{\delta}_r\hat{y}_{t-r} - \hat{\omega}_0 x_{t-b} - \cdots - \hat{\omega}_s x_{t-b-s} \qquad (11.2.15)$$

递推计算出 $\hat{y}_t = \hat{\delta}(B)\hat{\omega}(B)x_{t-b}$, 然后由 $\hat{n}_t = y_t - \hat{y}_t$ 计算噪声序列。在前面的每一种情形中，对 $\hat{n}_t$ 自相关函数的研究都有助于噪声模型的识别。

噪声的识别还可以利用经过白噪化后输入和输出的相关函数来进行，即按照下文所述方法。假设输入可以被完全白噪化而给出

$$\beta_t = v(B)\alpha_t + \epsilon_t \qquad (11.2.16)$$

其中，已知的关系式

$$\epsilon_t = \theta_x^{-1}(B)\phi_x(B)n_t \tag{11.2.17}$$

将 ϵ_t 和 n_t 联系起来。如果对 ϵ_t 可以找到一个随机模型，那么，利用 (11.2.17) 可以对 n_t —从而也是对 N_t 导出模型。如果我们记 $v(B)\alpha_t = u_t$, 便有 $\beta_t = u_t + \epsilon_t$, 同时，我们给出了有关 x_t 和 n_t 独立性的假设，从而满足了有关 u_t 和 ϵ_t 的独立性，我们可以写成

$$\gamma_{\beta\beta}(k) = \gamma_{uu}(k) + \gamma_{\epsilon\epsilon}(k) \tag{11.2.18}$$

因为 α_t 是白噪声，(3.1.8) 曾给出了线性过程的自相关函数，利用这一结果可以得出 $\gamma_{uu}(k)$ 。于是，由 (11.2.10) 有

$$\gamma_{uu}(k) = \sigma_\alpha^2 \sum_{j=0}^{\infty} v_j v_{j+k} = \frac{1}{\sigma_\alpha^2} \sum_{j=0}^{\infty} \gamma_{\alpha\beta}(j)\gamma_{\alpha\beta}(j+k)$$

因此，利用 (11.2.18)ϵ_t 的自协方差可由 $\gamma_{\epsilon\epsilon} = \gamma_{\beta\beta}(k) - \gamma_{uu}(k)$ 得出，相应的自相关为

$$\rho_{\epsilon\epsilon}(k) = \frac{\gamma_{\epsilon\epsilon}(k)}{\gamma_{\epsilon\epsilon}(0)} = \frac{\rho_{\beta\beta}(k) - \gamma_{uu}(k)/\gamma_{\beta\beta}(0)}{1 - \gamma_{uu}(0)/\gamma_{\beta\beta}(0)}$$

$$= \frac{\rho_{\beta\beta}(k) - \sum_{j=0}^{\infty}\rho_{\alpha\beta}(j)\rho_{\alpha\beta}(j+k)}{1 - \sum_{j=0}^{\infty}\rho_{\alpha\beta}^2(j)}$$

这样，在具体问题中，白噪化传递函数的估计就是必不可少的了。在得出近似的白噪化传递函数后，另外还有被传递的输入和输出之间互相关的估计值 $r_{\alpha\beta}(j)$, 以及被传递输出的自相关的估计值

表 11.2　煤气炉数据中噪声的自相关和偏相关函数

k	r_k	$\hat{\phi}_{kk}$	k	r_k	$\hat{\phi}_{kk}$
1	0.89	0.89	7	0.01	−0.2
2	0.71	−0.43	8	−0.03	0.01
3	0.51	−0.13	9	−0.05	−0.01
4	0.32	0.02	10	−0.04	0.08
5	0.17	0.04	11	−0.03	−0.06
5	0.07	−0.02	12	−0.03	−0.10

$r_{\beta\beta}(j)$, 把这些值代入便可得到 $\rho_{\epsilon\epsilon}(k)$ 的粗估计值.

在煤气炉例子中的应用 表 11.2 给出了噪声序列 $\hat{N}_t = Y_t - \hat{\mathcal{Y}}_t$ 的前 12 个样本自相关和偏相关函数值, 这里 $\hat{\mathcal{Y}}_t = \hat{\delta}^{-1}(B)\hat{\omega}(B)X_{t-3}$ 是用前面得到的传递函数的初步估计, 由 (11.2.15) 算出来的. 即其值由下式计算

$$\hat{\mathcal{Y}}_t = 0.57\hat{\mathcal{Y}}_{t-1} - (0.53X_{t-3} + 0.33X_{t-4} + 0.51X_{t-5})$$

$\hat{N}_t$ 的偏相关表明, 二阶的自回归模型可能是合适的描述, 从 $\hat{N}_t$ 的值对 AR(2) 模型的最小二乘估计得出

$$(1 - 1.54B + 0.64B^2)N_t = a_t \qquad (11.2.19)$$

其中 $\hat{\sigma}_a^2 = 0.057$.

因此由本节和 11.2.2 节的分析得出, 煤气炉模型识别为

$$Y_t = \frac{\omega_0 - \omega_1 B - \omega_2 B^2}{1 - \delta_1 B - \delta_2 B^2} X_{t-3} + \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t \qquad (11.2.20)$$

且初估计为 $\hat{\omega}_0 = -0.53, \hat{\omega}_1 = 0.33, \hat{\omega}_2 = 0.51, \hat{\delta}_1 = 0.57, \hat{\delta}_2 = 0.02, \hat{\phi}_1 = 1.54, \hat{\phi}_2 = -0.64$, 它们可以作为在 11.3 节中所述非线性估计方法的粗略的起始值.

11.2.4 传递函数模型识别的一般考虑

有关我们刚刚描述的传递函数模型和噪声模型的识别过程, 现在可以进行稍一般的考察.

1. 在许多实际情形中, 当噪声的影响显著时, 带有延迟的一阶或二阶系统 —如 (11.2.13) 所给出的或其某种简化形式 —常常能提供一个能被数据验证的精确的模型.

2. 有效的估计只有在假设模型形式为已知时才有可能得到. 因此, 由 (11.2.12) 所给出的估计 $\hat{v}_k$ 一般必定是低效的. 将它们

用于识别阶段，是因为它们便于计算，且能够告诉我们值得用更精细的方法去拟合的模型的形式.

3. 即使这些 $\hat{v}$ 是有效估计，为了充分刻划脉冲响应函数所需要的 $\hat{v}$ 的数目相当多，远比传递函数模型中参数的个数多. 有时，一个恰当的传递函数模型中 δ 和 ω 可以精确地估计出来，但是相应的 v 的估计值却有很大的方差，并且高度相关.

4. 估计量

$$r_{\alpha\beta}(k) = \hat{v}_k \frac{s_\alpha}{s_\beta}$$

的方差是与 $1/n$ 同阶的. 我们可以预料，估计值 $r_{\alpha\beta}(k)$ 以及 $\hat{v}_k$ 将会淹没在噪声中，除非 σ_α 与残差噪声相比足够大，或是 n 很大. 因此，识别过程要求输入 X 的变动与噪声引起的变动相比足够大，或者是可以获得大量数据. 这些要求煤气炉数据都能满足，如图 11.3 所示，其初步识别是非常成功的. 当这些要求得不到满足，识别过程就可能失败. 通常这意味着只能从现有数据得到很粗略的估计. 但是，几种基本的建模方法还是可行的，如设定某个相近而又简单的传递函数 / 噪声模型，直接用下一节的最小二乘法拟合，以及必要时使用诊断检验，使得模型更精细.

5. 或许应该强调一下，白噪化序列 α_t 和 β_t(特别是它们的互相关函数 $r_{\alpha\beta}(k)$) 只用于识别传递函数模型的形式这一目的. 一旦模型形式确定了，就要用原序列 X_t 和 Y_t(而不是白噪化序列) 去有效地估计模型参数，以及进行预测等等.

6. 另一种识别传递函数 - 噪声模型关系的方法由 Haugh 和 Box[216] 提出， Pristly[220 第 9 章] 也讨论了类似的思想. 这个一般方法可称为“双重线性化”，该方法涉及输入和输出两个序列的白噪化. 分别对输入和输出过程 (X_t 和 Y_t) 建立各自的单变量 ARIMA 模型，然后考察从这两个模型所得残差 (单变量白噪声) 的相关结构. 通过考察可以把联系两个残差序列的传递函数模型识别出来，从 X_t 和 Y_t 各自单变量的 ARIMA 模型中过滤 (即从观测序列到白噪声残差的滤波) 出的信息中，可以确定出用原序列

表示的模型。尽管这种方法有时是有用的，但是，在某些场合，该方法可能会使得所确定的传递函数模型的最终形式过于复杂，这是由于单变量白噪化系数存在的缘故。

模型唯一性的欠缺 假设一个具体的动态系统用下面模型表示

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + \varphi^{-1}(B)\theta(B)a_t \tag{11.2.21}$$

那么，也可等价地表示为

$$L(B)Y_t = L(B)\delta^{-1}(B)\omega(B)X_{t-b} + L(B)\varphi^{-1}(B)\theta(B)a_t \tag{11.2.22}$$

其中 $L(B)$ 是任意的公因子，从而是冗余的．类似于在节 7.3.5 中关于 ARMA 模型参数冗余性的讨论。为使 (11.2.21) 中模型参数化具有唯一性，这蕴含必须避免算子 $\delta(B)$ 和 $\omega(B)$ 或算子 $\varphi(B)$ 和 $\theta(B)$ 存在公因子的可能性。为了减少迭代涉及形式过于复杂模型的机会，我们应该把策略基于以下考虑：

1. 由于比较简单的传递函数模型 — 一阶或二阶，有延迟或无延迟 — 常常是恰当的，因此，反复建模步骤应从较简单的模型开始，并寻求更进一步简化的可能性，只有在非常必要时才转向更复杂的模型。

2. 不断地寻求机会，以便消去作用在 X_t,Y_t 和 a_t 上一个或多个算子的公因子．在实际中，我们所处理的是估计出来的系数，这会有较大的误差，所以只能期望近似的因子分解，这就需要发挥充分的想象去发现可能的因子分解．重新拟合、检验因子模型，以证明所作化简是否正确

3. 当因子化简的可能性存在但却被忽视时，最小二乘估计过程会变得极不稳定，这是因为在参数空间中最小值将在一条直线或一个曲面上，而不是一个点。反之，求解的不稳定也意味着存在模型简化的可能性。如前文中所强调的，在拟合模型之前应先进行识别，其原因之一就是在参数化中避免冗余，即实现参数的简约。

在附录 A11.1 中给出了另一种识别传递函数模型的方法，它可以很容易地推广到对多输入的处理。

注　如果在 (11.2.22) 中的算子 $L(B)$ 令其等于 $\varphi(B)\delta(B)$，我们将得到

$$\varphi(B)\delta(B)Y_t=\varphi(B)\omega(B)X_{t-b}+\delta(B)\theta(B)a_t \tag{11.2.23}$$

该式可写成 $\delta^*(B)Y_t=\omega^*(B)X_{t-b}+\theta^*(B)a_t$ 。形如 (11.2.23) 的一般形式的模型在经济学文献中 (如 Hannan 和 Deistler[101], Hannan,Dunsmuir,Deistler[102] 和 Reinsel[166]) 称为 ARMAX 模型。可以看到，在这种形式中，算子之间出现公共因子的可能性相当大。

11.3　传递函数模型的识别与拟合

11.3.1　条件平方和函数

对于尝试识别的模型

$$y_t=\delta^{-1}(B)\omega(B)x_{t-b}+n_t \tag{11.3.1}$$

其中 $y_t=\nabla^d Y_t,x_t=\nabla^d X_t,n_t=\nabla^d N_t$ 都是平稳过程，且

$$n_t=\phi^{-1}(B)\theta(B)a_t \tag{11.3.2}$$

我们现在来考虑上述模型中参数 $b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的同时有效的估计问题。假设有 $n=N-d$ 对值可用于分析，Y_t 和 X_t(若 $d>0$ 则为 y_t 和 x_t) 记关于期望值的离差。这些期望值可以和其它参数一起来估计，但是对于一个其长度值得去做分析的时间序列来说，通常使用样本均值作为估计值就够了。当 $d>0$ 时，y_t 和 x_t 的均值为零常常是成立的。如果序列开始之前的初始值 $\boldsymbol{x}_0,\boldsymbol{y}_0$ 和 $\boldsymbol{a}_0$ 可以得到，对于给出的数据及任意选定的参数 $(b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\theta})$ 和初始值 $(\boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{a}_0)$，我们可以对于 $t=1,2,\cdots,n$ 顺次计算出 $a_t=$

$\boldsymbol{a}_t(b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\theta}|\boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{a}_0)$．在 a 的正态假定下，通过对条件平方和函数

$$S_0(b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\theta})=\sum_{t=1}^{n}a_t^2(b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\theta}\mid\boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{a}_0) \qquad (11.3.3)$$

的极小化，可以得到参数极大似然估计很好的近似值。

计算 a 的三阶段方法　给定适当的初始值，对于任何具体选定的参数值都可以使用下面的三阶段方法来实现 a 的生成.

首先，传递函数模型的输出 y_t 可由下式计算

$$\widetilde{y}_t=\delta^{-1}(B)\omega(B)x_{t-b}$$

即

$$\delta(B)\widetilde{y}_t=\omega(B)x_{t-b}$$

或

$$\widetilde{y}_t-\delta_1\widetilde{y}_{t-1}-\cdots-\delta_r\widetilde{y}_{t-r}=\omega_0x_{t-b}-\omega_1x_{t-b-1}-\cdots-\omega_sx_{t-b-s} \qquad (11.3.4)$$

算出序列 $\widetilde{y}_t$ 后，利用 (11.3.1), 噪声 n_t 可由下式得到

$$n_t=y_t-\widetilde{y}_t \qquad (11.3.5)$$

最后， a 可有 (11.3.2) 用如下形式得出

$$a_t=\theta^{-1}(B)\phi(B)n_t$$

即

$$a_t=\theta_1a_{t-1}+\cdots+\theta_qa_{t-q}+n_t-\phi_1n_{t-1}-\cdots-\phi_pn_{t-p} \qquad (11.3.6)$$

初始值　正如节 7.3.1 中对随机模型估计的讨论一样，如果差分方程从某个 t 值开始，这时所有过去的 x 和 y 都是已知的了，瞬时状态的影响可以看作很小．于是， (11.3.4) 中的 $\widetilde{y}_t$ 从 $t=u_1$

开始向后计算，其中 u 是 r 和 $s+b$ 中较大者. 这意味着 n_{u+1} 之后的 n_t 可以得到，因此，如果未知的 a 设定等于其无条件均值 (零), 那么， a_t 可以从 a_{u+p+1} 开始向后计算. 这样，条件平方和函数为

$$S_0(b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{t=u+p+1}^{n} a_t^2(b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\theta} \mid \boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{a}_0) \qquad (11.3.7)$$

使用煤气炉数据的例子　对于这些数据，模型 (11.2.20) 即

$$Y_t = \frac{\omega_0 - \omega_1 B - \omega_2 B^2}{1 - \delta_1 B - \delta_2 B^2} X_{t-3} + \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t$$

已经识别出来。那么，方程 (11.3.4),(11.3.5) 和 (11.3.6) 成为

$$\mathcal{Y}_t = \delta_1 \mathcal{Y}_{t-1} + \delta_2 \mathcal{Y}_{t-2} + \omega_0 X_{t-3} - \omega_1 X_{t-4} - \omega_2 X_{t-5} \qquad (11.3.8)$$

$$N_t = Y_t - \mathcal{Y}_t \qquad (11.3.9)$$

$$a_t = N_t - \phi_1 N_{t-1} - \phi_2 N_{t-2} \qquad (11.3.10)$$

因此， (11.3.8) 可以用来生成 $t=6$ 以后的 $\mathcal{Y}_t$, 而 (11.3.10) 用来生成 $t=8$ 以后的 a_t . 对于足够长的序列来说，由此而引起的信息损失是不足道的. 例如，由于气炉数据 N =296, 序列开始时损失了 7 个值，这几乎不会有任何实际后果. 为了说明起见，表 11.3 中给出了煤气炉数据前几个 a_t 值的计算，其中

$$b=3\ , \quad \delta_1 = 0.1\ , \quad \delta_2 = 0.1\ , \quad \phi_1 = 0.1\ , \quad \phi_2 = 0.1$$
$$\omega_0 = 0.1\ , \quad \omega_1 = -0.1\ , \quad \omega_2 = -0.1$$

第二和第四列 X_t 和 Y_t 的值是用本书第五部分“时间序列汇集 ”中给出的序列值减去均值 $\overline{X} = -0.057$ 和 $\overline{Y} = 53.51$ 得到的.

表 11.3　对于煤气炉数据前几个 a 值的计算，其中 b=3, δ_1=0.1, δ_2=0.1, ω_0=−0.1, ω_1=− −0.1, ω_2=−0.1, ϕ_1=0.1, ϕ_2=0.1

t	X_t	$\mathcal{Y}_t$	Y_t	N_t	a_t
1	−0.052	—	0.29	—	—
2	0.057	—	0.09	—	—
3	0.235	—	−0.01	—	—
4	0.396	—	−0.01	—	—
5	0.430	—	−0.11	—	—
6	0.498	0.024	−0.41	−0.434	—
7	0.518	0.071	−0.81	−0.881	—
8	0.405	0.116	−1.11	−1.226	−1.094
9	0.184	0.151	−1.31	−1.461	−1.250
10	−0.123	0.171	−1.51	−1.681	−1.412

在前面我们假设了 $b=3$。为了估计 b, 可以对可能范围内的每个 b 值去计算使条件平方和极小的 $\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的值，这样就可以得到关于 $b,\boldsymbol{\delta},\boldsymbol{\omega},\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的总体最小值.

11.3.2 非线性估计

非线性最小二乘可以用来得到最小二乘估计及其近似标准差，这与 7.2.4 节中为了拟合随机模型所给出的方法很类似。当平方和函数大致接近于二次时，这种算法运行很好，但是，有时计算过程会出现问题，特别是参数高度相关时 (例如，模型由于很接近的公共因式分解而近似于奇异) 或者在有些情形下，估计值接近于允许参数空间的边界. 在出现困难时，可以选择参数空间中某两维的截面，画出平方和的等值线图，这样将使估计的情势明朗化。

这种算法可推出如下：在迭代的任一阶段，对于某个固定的延迟参数 b, 假定其余参数的最佳推测被得到，并记为

$$\boldsymbol{\beta}_0' = (\delta_{1,0},\cdots,\delta_{r,0};\omega_{0,0},\cdots,\omega_{s,0};\phi_{1,0},\cdots,\phi_{p,0};\theta_{1,0},\cdots,\theta_{q,0})$$

现在 $a_{t,0}$ 是用 11.3.1 中的方法从模型算出的值，对于推测参数值

$\boldsymbol{\beta}_0, a_t$ 关于这些参数的负导数表示如下:

$$
\begin{aligned}
d_{i,t}^{(\delta)} &= -\left.\frac{\partial a_t}{\partial \delta_i}\right|_{\boldsymbol{\beta}_0} \qquad & d_{j,t}^{(\omega)} &= -\left.\frac{\partial a_t}{\partial \omega_j}\right|_{\boldsymbol{\beta}_0} \\
d_{g,t}^{(\phi)} &= -\left.\frac{\partial a_t}{\partial \phi_g}\right|_{\boldsymbol{\beta}_0} \qquad & d_{h,t}^{(\theta)} &= -\left.\frac{\partial a_t}{\partial \theta_h}\right|_{\boldsymbol{\beta}_0}
\end{aligned} \tag{11.3.11}
$$

那么， $a_t = a_t(\boldsymbol{\beta})$ 关于参数值 $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ 的泰乐展式可重新整理为如下形式

$$
\begin{aligned}
a_{t,0} \sim & \sum_{i=1}^{r}(\delta_i - \delta_{i,0})d_{i,t}^{(\delta)} + \sum_{j=0}^{s}(\omega_j - \omega_{j,0})d_{j,t}^{(\omega)} \\
& + \sum_{g=1}^{p}(\phi_g - \phi_{g,0})d_{g,t}^{(\phi)} + \sum_{h=1}^{q}(\theta_h - \theta_{h,0})d_{h,t}^{(\theta)} + a_t
\end{aligned} \tag{11.3.12}
$$

用标准的线性最小二乘法来拟合这个线性化的方程，我们可以象节 7.2 中一样得出修正值 $\delta_i - \delta_{i,0}, \omega_j - \omega_{j,0}$, 等等. 把这些修正值加到第一次推测值 $\boldsymbol{\beta}_0$ 上，就生成了第二组推测值，重复这一过程直到收敛。

与随机模型一样 (见第 7 章，特别是 7.2.3 节), 导数可以递推地计算。但是，用标准的非线性最小二乘计算程序来工作是最简单的，其中的导数是数值方法确定的，而且为了避免不稳定 (见第 7 章), 选择 "有限制的迭代 " 是有效的。

如 7.2.2 节所述，估计值的协方差阵可以从矩阵 $(\boldsymbol{X}'_{\beta}X_{\beta})^{-1}\hat{\sigma}_a^2$ 的收敛值得到，而且已经证明，最小二乘估计 $\hat{\boldsymbol{\beta}}$ 具有多维正态渐近分布 (如见 Pierce[156],Reinsel[166]) 。如果需要估计 b(是整数), 可以对一系列的 b 进行迭代直到收敛，然后选择平方和最小的 b 值。最小二乘估计 $\hat{\boldsymbol{\beta}}$ 的协方差阵有一种特殊的特性 (见 Piercce[156]) , 即它将是近似的分块对角阵 , 对角线上的两块分别是参数 $(\hat{\delta}_1, \cdots, \hat{\delta}_r, \hat{\omega}_0, \cdots, \hat{\omega}_s)'$ 和 $(\hat{\phi}_1, \cdots \hat{\phi}_p, \hat{\theta}_1, \cdots, \hat{\theta}_q)'$ 的协方差阵. 因此，模型传递函数部分的参数估计值与模型噪声部分的估

计接近于不相关，这是由于在模型中假定了输入 X_t 和白噪声 a_t 之间是独立的。

与第 7 章中所讨论 ARMA 模型的情形一样 (例如可见 Newbold[148]), 更精确的平方和方法和极大似然方法也可用于估计传递函数 - 噪声模型 (11.3.1) 和 (11.3.2) 的估计. 特别是在处理模型的噪声部分时，对于给定的 $b,\boldsymbol{\delta}$ 和 $\boldsymbol{\omega}$, 噪声序列 n_t 可用 (11.3.5) 构造出，然后对于 $t = u + 1, \cdots, n$ 的 n_t 值可以值接应用关于 ARMA(p, q) 模型的精确平方和及似然方法。如节 7.5 所述，为计算精确似然值，状态空间模型 -Kalman 滤波方法也可以使用。但是，对于中等或较大的 n 及非季节数据，条件方法和精确方法之间差别不大。

11.3.3 用残差进行诊断检验

严重的模型失当可以通过下面的考察来检测:

1. 考察自相关函数 $r_{\hat{a}\hat{a}}(k)$, 它对应于所拟合模型的残差序列 $\hat{a}_t = a_t(\hat{b}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$.

2. 考察涉及输入和残差的某些互相关函数: 特别是白噪化输入 α_t 和残差 $\hat{z}_t$ 之间的互相关函数 $r_{\alpha\hat{a}}(k)$.

假定模型 (必要时在适当差分之后) 可写为

$$\begin{aligned} y_t &= \delta^{-1}(B)\omega(B)x_{t-b} + \phi^{-1}(B)\theta(B)a_t \\ &= v(B)x_t + \psi(B)a_t \end{aligned} \tag{11.3.13}$$

现在，假设我们选择了一个不恰当的模型，由此得出残差 a_{0t}, 这里

$$y_t = v_0(B)x_t + \psi_0(B)a_{0t}$$

那么

$$a_{0t} = \psi_0^{-1}(B)[v(B) - v_0(B)]x_t + \psi_0^{-1}(B)\psi(B)a_t \tag{11.3.14}$$

因此, 如果选择了错误的模型, 一般说来显然会有 a_{0t} 为自相关, 并且 a_{0t} 会与 x_t 以及生成 x_t 的 α_t 互相关.

现在考虑两种特殊情形： (1) 传递函数模型是正确的，但噪声模型不正确， (2) 传递函数模型不正确。

传递函数模型正确，噪声模型不正确　如果 $v_0(B) = v(B)$ 但 $\psi_0(B) \neq \psi(B)$, 那么 (11.3.14) 成为

$$a_{0t} = \psi_0^{-1}(B)\psi(B)a_t \tag{11.3.15}$$

因此， a_{0t} 将不会与 x_t 或 α_t 互相关。但是 a_{0t} 过程将是自相关的，且自相关函数的形式可以指出噪声结构应做的适当修改。

传递函数模型不正确　由 (11.3.14) 显然可知若传递函数不正确，不仅 a_{0t} 会与 x_t(及 α_t) 互相关，而且 a_{0t} 会自相关。甚至在噪声模型是正确的时候仍然是这样，因为这时 (11.3.14) 将成为

$$a_{0t} = \psi^{-1}(B)[v(B) - v_0(B)]x_t + a_t \tag{11.3.16}$$

不管噪声模型正确与否，互相关分析都会指出传递函数模型所需要的修改。通过考虑白噪化后的模型，这一点便十分清楚。如果假定输出和输入都是经变换的，从而输入是白噪声，那么，正如 (11.2.8), 我们可以将模型写为

$$\beta_t = v(B)\alpha_t + \epsilon_t$$

其中 $\beta_t = \theta_x^{-1}(B)\phi_x(B)y_t$, $\epsilon_t = \theta_x^{-1}(B)\phi_x(B)n_t$ 。现在考虑量值

$$\epsilon_{0t} = \beta_t - v_0(B)\alpha_t$$

由于 $\epsilon_{0t} = [v(B) - v_0(B)]\alpha_t + \epsilon_t$ 正如节 11.2.1 所述， ϵ_{0t} 和 α_t 的互相关度量了正确和不正确的脉冲响应函数之间的偏差。特别地，如在 (11.2.11) 中，

$$v_k - v_{0k} = \frac{\rho_{\alpha\epsilon_0}(k)\sigma_{\epsilon_0}}{\sigma_\alpha} \qquad (k = 0, 1, 2, \cdots) \tag{11.3.17}$$

11.3.4 用于残差的具体检验

在实际中，我们并不能准确地知道过程的参数，但是必须将我们的检验运用于最小二乘拟合后的残差 $\hat{a}_t$ 。即使拟合模型的函数形式是恰当的，参数估计值也会与真值有所不同，而且残差 $\hat{a}_t$ 相关的分布也在一定程度上与 a_t 的自相关不同。因此，在利用前节的结果去推测残差相关的特性时，谨慎行事是很必要的。下文的简要讨论是基于 [157] 中给出的更详细的研究。

自相关检验　假设一个传递函数 - 噪声模型已经用最小二乘法拟合出来，并且将参数用其最小二乘估计代替后算出了残差 $\hat{a}_t$，这些残差的自相关函数估计值 $r_{\hat{a}\hat{a}}(k)$ 也算出。那么，正如我们已经看到的：

1. 如果自相关函数 $r_{\hat{a}\hat{a}}(k)$ 表现出显著的相关，则意味着模型不恰当。

2. 如果互相关检验并没有显示出传递函数模型不恰当，则不恰当可能产生于拟合噪声模型 $n_t = \psi_0(B)a_t$ 。

在后一种情况，识别一个辅助模型

$$\hat{a}_{0t} = T(B)a_t$$

用以描述主模型残差的相关，由 (11.3.15) 可大致表明，下式

$$n_t = \psi_0(B)T(B)a_t$$

能够作为修改后的模型噪声。当我们判断自相关函数的估计值是否明显偏离于零，以此说明是否有非零的理论值，这时，类似于节 8.2.1 中所讨论的一些事实必须牢记在心。

在扣除初始值后，实际上有 $m = n - u - p$ 个值可用于计算。如果模型的函数形式是正确的，且用参数的真值代入，残差将会是白噪声自相关函数估计值并相互独立地以方差 $1/m$ 分布在零附近。如果用估计值来代替参数值，低滞后点上的相关分布特征会受到影响。具体地说，这些低滞后点上相关估计值的方差将会

比 $1/m$ 小得多，而这些值之间却可能是高度相关的。这样，k 很小时，把自相关估计值 $r_{\hat{a}\hat{a}}(k)$ 与“标准差” $1/\sqrt{m}$ 来比较将会极大地忽视其重要性。另外，在低滞后点上自相关函数的波动可能只是由于这些估计值之间的高度相关而诱发产生的。如果在这些低滞后点上的波动幅度与 $1\sqrt{m}$ 相比较小，那么它们可能只不过是偶然发生的，并不一定表示理论自相关的某种真实形式。

下面给出一个有用的总体检验，用来考虑因拟合而产生的分布影响。考虑前 K 个自相关函数估计值 $r_{\hat{a}\hat{a}}(1),\cdots,r_{\hat{a}\hat{a}}(K)$, 取 K 足够大，以使得模型写为 $y_t=v(B)+\psi(B)a_t$ 时，权数 ψ_j 当 $j>K$ 时可望被忽略。那么，如果模型的形式是恰当的，量值

$$Q=m\sum_{k=1}^{K}r_{\hat{a}\hat{a}}^2(k) \tag{11.3.18}$$

近似服从自由度为 $K-p-q$ 的 χ^2 分布。注意，χ^2 的自由度取决于噪声模型中参数的个数，而不是传递函数模型中参数的个数。将 Q 与 χ^2 的百分位点表做比较，我们可以得到关于模型恰当性假设的近似检验。但是，在实际中推荐使用的统计量不是 (11.3.18)，而是修正后的统计量 $\widetilde{Q}=m(m+2)\sum\limits_{k=1}^{K}(m-k)^{-1}r_{\hat{a}\hat{a}}^2(k)$, 这同节 8.2.2 中对 ARMA 模型所定义的 (8.2.3) 一样。这是因为在模型恰当的原假设下 $\widetilde{Q}$ 比 Q 更接近于 χ^2 分布。

互相关检验　正如我们在 11.3.3 节中所看到的：

1. 互相关 $r_{x\hat{a}}(k)$ 显著非零的特性意味着传递函数模型的不当。

2. 一种稍许不同的互相关分析可以指出需做何种修改。具体地，如果拟合的传递函数是 $\hat{v}_0(B)$, 我们考虑 $\hat{\epsilon}_{0t}=\beta_t-\hat{v}_0(B)\alpha_t$ 和 α_t 这两个量之间的互相关，则描述差异程度 u_k-u_{0k} 的粗略估计量由下式给出

$$\frac{r_{\alpha\hat{\epsilon}_0}(k)S_{\hat{\epsilon}_0}}{S_\alpha}$$

假设模型具有正确的函数形式，并把真实的参数值代入。残差将是一个与 x 无关的白噪声，利用 (11.1.12) 可知，对一个有效长度为 m 的序列来说， $r_{xa}(k)$ 的方差近似为 $1/m$ 。但是与自相关 $r_{aa}(k)$ 不同，这些互相关不会大致不相关。一般地，如果 x 是自相关的，那么，互相关 $r_{xa}(k)$ 也是自相关的。事实上，正如在 (11.1.12) 中所见，在 x 和 a 不存在互相关的假设之下， $r_{xa}(k)$ 和 $r_{xa}(k+l)$ 之间的互相关系数是

$$\rho[r_{xa}(k), r_{xa}(k+l] \sim \rho_{xx}(l) \tag{11.3.19}$$

也就是说，互相关的自相关函数与原输入序列 x_t 的自相关函数是一样的。因此，若 x_t 是自相关的，一个完全恰当的传递函数模型会得出增长的互相关 $r_{x\hat{a}}(k)$，尽管幅度可能很小，但这种特征是明显的。如果进行检验时用白噪化输入 α_t 来计算互相关 $r_{\alpha\hat{a}}(k)$，那么上述影响会被消除。

和自相关一样，当参数值用估计值来替代时，自相关函数的分布特性会受到影响。但是，与自相关检验类似，可以根据互相关的大小得到关于模型恰当性假设的粗略的整体检验。为了利用这个检验，需要计算白噪化式输入 α_t 和残差 $\hat{z}_t$ 之间互相关 $r_{\alpha\hat{a}}(k)$ 的估计值，这里 $k=0,1,2,\cdots,K$，且 K 要选得足够大，以使 $j>K$ 以后 (11.3.13) 中的权数 v_j 和 ψ_j 可望忽略不计。与前面一样，因使用参数估计值来计算残差所造成的影响主要局限在低阶的互相关上，其方差大大低于 $1/m$，即使输入为白噪声时仍会高度相关。

然而，下述事实是成立的 (见 [157]):

$$S = m\sum_{k=0}^{K} r_{\alpha\hat{a}}(k) \tag{11.3.20}$$

近似地是自由度为 $K+1-(r+s+1)$ 的 χ^2 分布，这里 $(r+s+1)$ 是在传递函数模型中所拟合参数的个数。注意，自由度的数目和

噪声模型中拟合参数的个数是无关的。根据第 8 章中讨论的对 Q 统计量特性的研究，修正统计量 $\widetilde{S}=m(m+2)\sum_{k=0}^{K}(m-k)^{-1}r_{\alpha\hat{a}}(k)$ 可以用来代替 (11.3.20) 中的 S, 这是因为在零模型的假设下该统计量更准确地接近于 χ^2 分布，尽管对其特性的详细研究尚未作出 (然而还是可以参见 Poskitt 和 Tremayne[219] 的经验结论)。

11.4 拟合及检验传递函数模型的一些例子

11.4.1 煤气炉模型的拟合及检验

我们现在详细地说明将 11.3 节所述方法用于拟合模型

$$Y_t=\frac{\omega_0-\omega_1B-\omega_2B^2}{1-\delta_1B-\delta_2B^2}X_{t-3}+\frac{1}{1-\phi_1B-\phi_2B^2}a_t$$

这是在 11.2.2 和 11.2.3 中对煤气炉数据识别出来的模型。

非线性估计　利用在 11.2.2 节中推出的初始估计 $\hat{\omega}_0=-0.53$, $\hat{\omega}_1=0.33, \hat{\omega}_2=0.51, \hat{\delta}_1=0.57, \hat{\delta}_2=0.02, \hat{\phi}_1=1.54, \hat{\phi}_2=-0.64$, 并运用节 11.3.2 中所述条件最小二乘算法，经四次迭代可得到两位小数精度的最小二乘估计值. 为了检验在不利情形下迭代是否收敛，表 11.4 中列出了所有初始值取为 +0.1 或 –0.1 时迭代的过程. 即使如此，模型中所有 7 个参数都在 10 步之内迭代收敛，这个事实是令人鼓舞的.

表 11.4 的最后一行列出了 11.2.2 和 11.2.3 节中识别阶段得到的粗略的初步估计. 可以看到，在这个例子中它们与前一行给出的最小二乘估计很一致. 这样，最终拟合的传递函数模型是

$$\underset{(\pm0.21)(\pm0.14)}{(1-0.57B-0.01B^2)}Y_t=-\underset{(\pm0.08)(\pm0.15)(\pm0.16)}{(0.53+0.37B+0.51B^2)}X_{t-3} \tag{11.4.1}$$

拟合的噪声模型是

$$\underset{(\pm0.05)(\pm0.05)}{(1-1.53B+0.63B^2)}N_t=a_t \tag{11.4.2}$$

表 11.4 煤气炉数据非线性最小二乘拟合的收敛性

迭代次数	ω_0	ω_1	ω_2	δ_1	δ_2	ϕ_1	ϕ_2	平方和
0	0.10	−0.10	−0.10	0.10	0.10	0.10	0.10	13,601
1	−0.46	0.63	0.60	0.14	0.27	1.33	−0.27	273.1
2	−0.52	0.45	0.31	0.40	0.52	1.37	−0.43	92.5
3	−0.63	0.60	0.01	0.12	0.73	1.70	−0.76	31.8
4	−0.54	0.50	0.29	0.24	0.42	1.70	−0.81	19.7
5	−0.50	0.31	0.51	0.63	0.09	1.56	−0.68	16.84
6	−0.53	0.38	0.53	0.54	0.01	1.54	−0.64	16.60
7	−0.53	0.37	0.51	0.56	0.01	1.53	−0.63	16.60
8	−0.53	0.37	0.51	0.56	0.01	1.53	−0.63	16.60
9	−0.53	0.37	0.51	0.57	0.01	1.53	−0.63	16.60
初估计	−0.53	0.33	0.51	0.57	0.02	1.54	−0.64	

且有 $\hat{\sigma}_a^2 = 0.0561$, 这里括号中的范围是由非线性估计过程得出的 ±1 倍的标准差范围。

诊断检验 在把上述模型当作系统的恰当表示而接受之前，应该采用 11.3.4 节所述的自相关和互相关检验。表 11.5 列出了前 36 步滞后的残差自相关，以及假设模型为恰当时它们的标准差上限 $1/\sqrt{m}(m = 289)$ 。从单个自相关的特性来看似乎没有模型不恰当的证据，但是通过 $\widetilde{Q}$ 判别值的计算得到了证实，即

$$\widetilde{Q} = (289)(291)\sum_{k=1}^{36}(289-k)^{-1}r_{\alpha\hat{a}}^2(k) = 43.8$$

把 $\widetilde{Q}$ 与自由度为 $K-p-q=36-2-0=34$ 的 χ^2 相比较，并没有提供对模型恰当性存疑的依据。

表 11.6(a) 给出了输入和残差之间互相关函数 $r_{x\hat{a}}(k)$ 的前 36 步滞后的值，以及它们标准差的上限 $1/\sqrt{m}$ 。可以看到，尽管互相关 $r_{x\hat{a}}(k)$ 与其标准差上限相比不是特别大，但它们本身是高度相关的。这正是预料之中的，因为如 (11.3.19) 所示，该互相关与输入 x_t 服从同一随机过程，而我们已经看到在这个例子中输入是高度自相关的。

表 11.5　煤气炉模型拟合残差的自相关函数 $r_{\hat{a}\hat{a}}(k)$ 估计值

滞后 k	$r_{\hat{a}\hat{a}}(k)$												标准差上限
1–12	0.02	0.06	−0.07	−0.05	−0.05	0.12	0.03	0.03	−0.08	0.05	0.02	0.10	±0.06
13–24	−0.04	0.05	−0.09	−0.01	−0.08	0.00	−0.12	0.00	−0.01	0.08	0.02	−0.01	±0.06
25–36	0.04	−0.02	0.02	0.09	−0.12	0.06	−0.03	−0.06	0.11	0.02	0.03	0.06	±0.06

表 11.6　煤气炉数据互相关函数估计值

滞后 k													标准差上限
	(a) 输入和输出残差之间的 $r_{x\hat{a}}(k)$												
1–12	0.00	0.00	0.00	0.00	0.00	0.00	−0.01	−0.02	−0.03	−0.05	−0.06	−0.05	±0.06
13–24	−0.03	−0.03	−0.03	−0.07	−0.10	−0.12	−0.12	−0.10	−0.04	−0.01	−0.01	−0.02	±0.06
25–36	−0.03	−0.04	−0.04	−0.02	−0.01	0.02	0.04	0.05	0.06	0.07	0.07	0.06	±0.06
	(b) 白噪化输入和输出残差之间的 $r_{\alpha\hat{a}}(k)$												
1–12	−0.06	0.03	−0.01	0.00	0.01	0.01	0.01	−0.04	0.02	0.07	−0.03	−0.02	±0.06
13–24	−0.03	−0.11	0.02	0.04	0.04	0.01	0.01	−0.15	−0.03	−0.07	−0.08	0.02	±0.06
25–36	−0.01	0.02	0.05	−0.07	0.00	0.04	−0.15	0.04	0.03	−0.02	0.00	−0.03	±0.06

表 11.6(b) 中给出了相应的 $\hat{a}_t$ 与白噪化输入 α_t 之间的互相关。$\widetilde{S}$ 判别量为

$$\widetilde{S}=(289)(291)\sum_{k=0}^{35}(289-k)^{-1}r_{\alpha\hat{a}}^2(k)=32.1$$

将 $\widetilde{S}$ 与自由度 $K+1-(r+s+1)=36-5=31$ 的 χ^2 分布表作比较，仍然不能提供模型不当的依据。

阶跃和脉冲响应　(11.4.1) 中的估计值 $\hat{\delta}_2=0.01$，它与标准差 ±0.14 相比非常小，事实上，可以将参数 $\hat{\delta}_2$ 从模型中略去，而不影响剩余参数的估计精度。煤气炉数据的传递函数 – 噪声联合模型的最终形式是

$$Y_t=\frac{-(0.53+0.37B+0.51B^2)}{1-0.57B}X_{t-3}+\frac{1}{1-1.53B+0.63B^2}a_t$$

相应于传递函数

$$(1-0.57B)Y_t=-(0.53+0.37B+0.51B^2)X_{t-3}$$

的阶跃和脉冲响应函数在图 11.6 中给出。利用 (10.2.5) 式，煤气炉数据的稳态增益是

$$g=\frac{-(0.53+0.37+0.51)}{1-0.57}=-3.3$$

这一结果与用互谱分析所得的结果 [122] 极为一致。

采样间隔的选择　在有可能进行选择时，与对系统预期的时间常数相比，采样间隔应取得相当短。在没有把握的情形下可重复试验几个采样间隔来进行分析。在选择采样间隔时，输出噪声是重要的，其方差应随间隔缩短而达到最小值。在煤气炉例子中，使用笔式记录仪来提供输入和输出的连续记录。我们实际分析的离散数据是从连续记录中在每隔 9 秒的离散点上读出来的。

因为从图 11.1 中轨迹的考察表明，选择这一间隔足以使输入和输出中发生的所有变化（除了记录笔轻微的振动）考虑进来。这类合理想法的采用在选择间隔时常常是可靠的依据。表 11.7 列

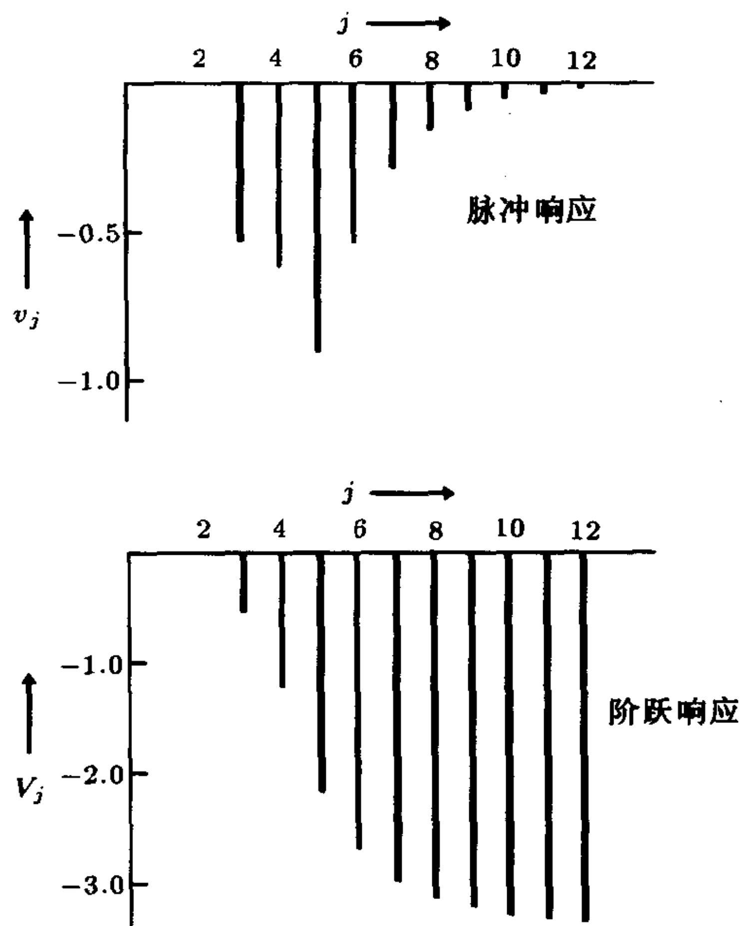


图 11.6 拟合煤气炉数据传递函数模型 $(1-0.57B)Y_t=-(0.53+0.37B+0.51B^2)X_{t-3}$ 的脉冲和阶跃响应

出了各种时间间隔之下对煤气炉数据估计出的均方误差 (由 $\sum(Y-\hat{Y})^2$ 除以适当的自由度个数得到，其中 $\hat{Y}$ 是拟合值) 。这些值也在图 11.7 中绘出。均方误差的变化甚微，直到采样间隔大约为 40 秒时才呈现迅速地上升。对于 9 秒、 18 秒和 27 秒的间隔，均方误差及实际绘出的阶跃响应几乎没有差别．但在使用 36 秒的间隔时，发生了相当大的变化。可见我们在本例中所用的 9 秒间隔采样实际上是保守的。

表 11.7 选择各种采样间隔时输出的均方误差

	间隔 (秒)						
	9	18	27	36	45	54	72
数据点数 N	296	148	98	74	59	49	37
M.S 误差	0.71	0.78	0.74	0.95	0.97	1.56	7.11

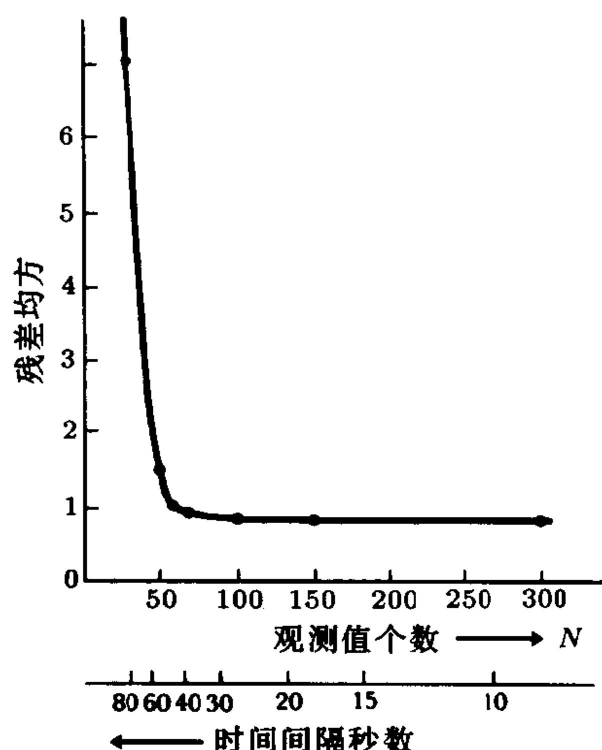


图 11.7 选择各种采样间隔时输出的均方误差

11.4.2 两输入的模拟例子

当输入序列超过一个时，模型的拟合没有任何实质性困难，只是需要处理的参数个数增加了。例如，对于两输入我们可以把模型写为

$$y_t = \delta_1^{-1}(B)\omega_1(B)x_{1,t-b_1} + \delta_2^{-1}(B)\omega_2(B)x_{2,t-b_2} + n_t$$

其中

$$n_t = \phi^{-1}(B)\theta(B)a_t$$

这里 $y_t = \nabla^d Y_t, x_{1,t} = \nabla^d X_{1,t}, x_{2,t} = \nabla^d X_{2,t}$, 及 $n_t = \nabla^d N_t$ 是平稳过程。为了计算 a_t, 我们先对参数 $b_1, \boldsymbol{\delta}_1, \boldsymbol{\omega}_1$ 的特定值计算

$$\widetilde{y}_{1,t} = \delta_1^{-1}(B)\omega_1(B)x_{1,t-b_1} \tag{11.4.3}$$

对 $b_2, \boldsymbol{\delta}_2, \boldsymbol{\omega}_2$ 的特定值计算

$$\widetilde{y}_{2,t} = \delta_2^{-1}(B)\omega_2(B)x_{2,t-b_2} \tag{11.4.4}$$

于是噪声 n_t 可由下式计算

$$n_t = y_t - \widetilde{y}_{1,t} - \widetilde{y}_{2,t} \tag{11.4.5}$$

最后可得

$$a_t = \theta^{-1}(B)\phi(B)n_t \tag{11.4.6}$$

模拟示例 显然，即使在简单的情形下也会导致数量很多的参数的估计. 下面的例子是两个输入变量和一阶延迟的模型，有 8 个未知参数。为了确定 11.3.2 节中所述的迭代非线性最小二乘算法是否可以用于获得该模型参数的估计，我们使用人工生成的数据进行一个试验，详见 [51] 。数据是由用 ∇ 表示的如下模型产生的:

$$Y_t = \beta + g_1\frac{1+\eta_1\nabla}{1+\xi_1\nabla}X_{1,t-1} + g_2\frac{1+\eta_2\nabla}{1+\xi_2\nabla}X_{2,t-1} + \frac{1}{1-\phi_1 B}a_t \tag{11.4.7}$$

其中 $\beta = 60, g_1 = 13.0, \eta_1 = -0.6, \xi_1 = 4.0, g_2 = -5.5, \eta_2 = -0.6, \xi_2 = 4.0, \phi_1 = 0.5$, 以及 $\sigma_a^2 = 9.0$ 。输入变量 X_1 和 X_2 是按照三次重复的随机 2^2 因子设计而变化的，假设每个输入条

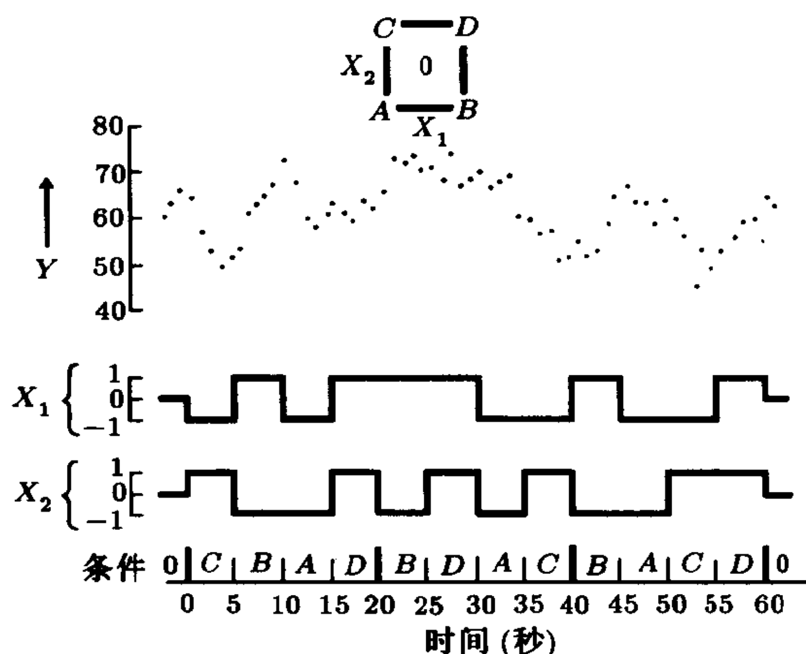


图 11.8 模拟两输入示例的数据 (序列 K)

件持续 5 分钟保持不变，而输出的观测每一分钟取一次。这些数据在图 11.8 中绘出，并在第五部分“时间序列汇总”中作为 K 序列给出。

利用第 7 章所述的有约束迭代非线性最小二乘算法得出最小二乘估计，只需要去计算 a 了。因此，对于参数 $g_1, g_2, \xi_1, \xi_2, \eta_1, \eta_2$ 的具体值，$\mathcal{Y}_{1,t}$ 和 $\mathcal{Y}_{2,t}$ 可由

$$(1+\xi_1\nabla)\mathcal{Y}_{1,t} = g_1(1+\eta_1\nabla)X_{1,t-1}$$
$$(1+\xi_2\nabla)\mathcal{Y}_{2,t} = g_2(1+\eta_2\nabla)X_{2,t-1}$$

得到，并用于计算

$$N_t = Y_t - \mathcal{Y}_{1,t} - \mathcal{Y}_{2,t}$$

最后，对于特定的参数 ϕ_1, a_t 可由下式算出

$$a_t = N_t - \phi_1 N_t$$

假设在试验开始之前，输入过程已经在其标准状态持续了一段时间，这样，$\mathcal{Y}_{1,t}, \mathcal{Y}_{2,t}$，从而 N_t 可以从 $t=0$ 开始计算，而 a_t 可从 $t=1$ 开始计算。

表 11.8　两输入情形，使用推测始值所做非线性最小二乘过程的收敛性

迭代次数	β	g_1	g_2	η_1	η_2	ξ_1	ξ_2	ϕ_1	平方和
0	59.19	10.00	−7.00	−0.50	−0.50	1.00	1.00	0.01	2,046.8
1	59.12	9.07	−6.37	−0.58	−0.56	1.33	1.31	0.24	1,085.4
2	59.24	8.38	−5.35	−0.70	−0.59	2.30	1.75	0.39	621.5
3	59.35	9.24	−3.98	−0.75	−0.55	3.45	1.95	0.36	503.5
4	59.41	11.90	−3.40	−0.75	−0.56	5.21	1.66	0.22	463.7
5	59.39	12.03	−3.52	−0.80	−0.57	4.99	1.76	0.21	461.8
6	59.39	12.08	−3.53	−0.79	−0.56	5.03	1.77	0.21	461.8
7	59.39	12.07	−3.53	−0.79	−0.56	5.03	1.77	0.21	461.8
8	59.39	12.07	−3.53	−0.79	−0.56	5.03	1.77	0.21	461.8

表 11.9 两输入情形，使用极端初值所做非线性最小二乘过程的收敛性

迭代次数	β	g_1	g_2	η_1	η_2	ξ_1	ξ_2	ϕ_1	平方和
0	59.19	0.10	0.10	0.10	0.10	0.10	0.10	0.10	2,496.8
1	59.19	0.24	−0.07	−1.57	0.48	1.77	−0.28	0.15	2,190.5
2	59.22	1.62	−0.29	−2.09	−2.24	−0.07	0.26	0.29	1,473.6
3	59.21	1.80	−0.77	−1.75	0.58	0.20	−0.10	0.56	1,016.8
4	59.21	3.01	−1.31	−1.15	−0.83	0.91	0.22	0.72	743.1
5	59.31	6.17	−2.82	−0.93	−0.65	3.03	1.20	0.67	611.4
6	59.61	15.83	−3.25	−0.70	−0.66	8.88	1.64	0.26	534.2
7	59.47	10.31	−3.48	−0.74	−0.56	3.52	1.63	0.23	501.9
8	59.41	11.89	−3.41	−0.74	−0.58	5.01	1.65	0.20	462.8
9	59.39	12.07	−3.52	−0.79	−0.57	5.04	1.76	0.21	461.8
10	59.39	12.07	−3.53	−0.79	−0.56	5.03	1.77	0.21	461.8
11	59.39	12.07	−3.53	−0.79	−0.56	5.03	1.77	0.21	461.8

使用两组不同的初值，对非线性最小二乘过程运行了两次。在第一组中，参数的选择表明一个对于该过程相当熟悉的人对参数初值可能做的猜测。在第二组中，β 的起始值选为所有观测值的均值 $\overline{Y}$，其它所有起始值都被置为 0.1. 因此，第二轮运行表示出一种非常极端的情形，实际中一般不会发生。表 11.8 列出了第一组初值，收敛在第五次迭代后出现，表 11.9 列出的是第二组初值，收敛在第九次迭代后出现。这些结果表明，在实际情形中对多重输入的处理在估计上没有太大的困难。

11.5 使用领先指标的预测

通常，在利用了某个相关序列 $X_t, X_{t-1}, \cdots$ 的信息后，对于时间序列 $Y_t, Y_{t-1}, \cdots$ 的预测可以得到相当程度的改善。特别是当 Y 的变化可以由 X 的变化预见时更是如此，这时经济学家称 X 为 Y 的“领先指标”。为了利用来自 Y 和 X 的信息得到最优预测，我们首先用已给出的形式建立一个传递函数 - 噪声模型将 Y 和 X 联系起来。采用前面的符号，假设恰当的模型是

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + \varphi^{-1}(B)\theta(B)a_t \qquad (b > 0) \qquad (11.5.1)$$

一般，该模型的噪声分量假设为与输入 X_t 统计独立，且是非平稳的，有

$$\varphi(B) = \phi(B)\nabla^d$$

那么如果

$$\nabla^d Y_t = y_t \quad 及 \quad \nabla^d X_t = x_t$$

$$y_t = \delta^{-1}(B)\omega(B)x_{t-b} + \phi^{-1}(B)\theta(B)a_t$$

我们对领先指标也要假设一个恰当的随机模型，设其为

$$X_t = \varphi_x^{-1}(B)\theta_x(B)\alpha_t \tag{11.5.2}$$

从而有

$$\varphi_x(B) = \phi_x(B)\nabla^d$$

$$x_t = \phi_x^{-1}(B)\theta_x(B)\alpha_t$$

11.5.1 最小均方误差预测

现在 (11.5.1) 可写作

$$Y_t = \nu(B)\alpha_t + \psi(B)a_t \tag{11.5.3}$$

其中 a 和 α 统计独立，且 $\nu(B) = \delta^{-1}(B)\omega(B)B^b\varphi_x^{-1}(B)\theta_x(B)$ 。与节 5.1.1 的论述相同，假设在原点 t 所作 Y_{t+l} 的预测 $\hat{Y}_t(l)$，其形式为

$$\hat{Y}_t(l) = \sum_{j=0}^{\infty}\nu_{l+j}^0\alpha_{t-j} + \sum_{j=0}^{\infty}\psi_{l+j}^0 a_{t-j}$$

那么

$$Y_{t+l} - \hat{Y}_t(l) = \sum_{i=0}^{l-1}(\nu_i\alpha_{t+l-i} + \psi_i a_{t+l-i}) + \sum_{j=0}^{\infty}[(\nu_{l+j} - \nu_{l+j}^0)\alpha_{t-j} + (\psi_{l+j} - \psi_{l+j}^0)a_{t-j}]$$

及

$$E[Y_{t+l}-\hat{Y}_t(l)]^2 =(\nu_0^2+\nu_1^2+\cdots+\nu_{l-1}^2)\sigma_\alpha^2+(1+\psi_l^2+\cdots+\psi_{l-1}^2)\sigma_a^2 + \sum_{j=0}^{\infty}[(\nu_{l+j}-\nu_{l+j}^0)^2\sigma_\alpha^2+(\psi_{l+j}-\psi_{l+j}^0)^2\sigma_a^2]$$

该式仅当 $\nu_{l+j}^0 = \nu_{l+j}$ 和 $\psi_{l+j}^0 = \psi_{l+j}$ 时最小。因此，在原点 t 处 Y_{t+l} 的最小均方误差预测 $\hat{Y}_t(l)$ 由 Y_{t+l} 在时刻 t 的条件期望给出。在理论上，该期望是以无穷远的过去直到当前原点 t 的信息序列作为条件的。如第五章中所述，这一结果是有实际作用的，因为通常预报值只是最明显地依赖于 X 和 Y 新近的过去值。

预测值的计算　现在 (11.5.1) 可写为

$$\varphi(B)\delta(B)Y_t = \varphi(B)\omega(B)X_{t-b}+\delta(B)\theta(B)a_t$$

我们将其记成

$$\delta^*(B)Y_t = \omega^*(B)X_{t-b}+\theta^*(B)a_t$$

于是，用方括号表示时刻 t 的条件期望，并记 $p^*=p+d$, 对于领先 l 步的预测，我们有

$$\begin{aligned}\hat{Y}_t(l)=[Y_{t+l}] =&\delta_1^*[Y_{t+l-1}]+\cdots+\delta_{p^*+r}^*[Y_{t+l-p^*-r}]+\omega_0^*[X_{t+l-b}]\\ &-\cdots-\omega_{p^*+s}j[X_{t+l-b-p^*-s}]+[a_{t+l}]-\theta_1^*[a_{t+l-1}]\\ &-\cdots-\theta_{q+r}^*[a_{t+l-q-r}]\end{aligned} \tag{11.5.4}$$

其中

$$[Y_{t+j}]=\begin{cases}Y_{t+j} & , \quad j\leqslant 0\\ \hat{Y}_t(j) & , \quad j>0\end{cases}$$

$$[X_{t+j}]=\begin{cases}X_{t+j} & , \quad j\leqslant 0\\ \hat{X}_t(j) & , \quad j>0\end{cases} \tag{11.5.5}$$

$$[a_{t+j}]=\begin{cases}a_{t+j} & , \quad j\leqslant 0\\ 0 & , \quad j>0\end{cases}$$

a_t 由 (11.5.1) 算出，若 $b > 1$, 该式等价于

$$a_t = Y_t - \hat{Y}_{t-1}(1)$$

因此，通过适当的替换，最小均方误差预测可很容易地直接用 (11.5.4) 和 (11.5.5) 算出。预测值 $\hat{X}_t(j)$ 用通常的方法 (见 5.2 节) 由模型 (11.5.2) 得到。

注意，(11.5.4) 和 (11.5.5) 中的条件期望是关于 Y_t 和 X_t 两个序列直到 t 时刻的值取得的，这一点很重要，但是，由于在 (11.5.1) 中假定了输入 X_t 和噪声 N_t 的独立性，因此，我们特别有

$$\hat{X}_t(j) = E[X_{t+j}|X_t, X_{t-1}, \cdots, Y_t, Y_{t-1}, \cdots] = E[X_{t+j}|X_t, X_{t-1}, \cdots]$$

也就是说，若给出了序列 X_t 的过去值，其未来值的最优预测只依赖于过去的 X, 而并不能被过去 Y 的补充信息所改善，因此 $\hat{X}_t(j)$ 的最优值可以直接由单变量模型 (11.5.3) 得到。

预测的方差　在

$$\delta(B)\varphi_x(B)\nu(B) = \omega(B)\theta_x(B)B^b$$

和

$$\varphi(B)\psi(B) = \theta(B)$$

中，令系数相等，显然可以得出权数 ν 和权数 ψ 。

领先 l 步预测误差的方差由下式给出

$$V(l) = E[Y_{t+l} - \hat{Y}_t(l)]^2 = \sigma_a^2 \sum_{j=b}^{l-1} \nu_j^2 + \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2 \qquad (11.5.6)$$

作为以往观测值加权组合的预测　对于任何给出的例子，仔细地考虑如何利用 X 和 Y 的既往值去预测 Y 的未来值，这是很具有启发性的。在 5.3.3 节中我们已经看到，怎样把预测值写成

序列既往值的线性组合。这样，对于领先指标的预测，我们可写成

$$\hat{X}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} X_{t-j+1} \tag{11.5.7}$$

若对于权数有 $\pi_j^{(l)} = \pi_j$, 模型 (11.5.2) 可写为下面形式

$$\alpha_t = X_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \cdots$$

在下式中

$$\varphi_x(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots)\theta_x(B)$$

并利用 (5.3.9), 令系数相等可得到

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{h=1}^{l-1} \pi_h \pi_j^{(l-h)} \tag{11.5.8}$$

用类似的方法，我们可将传递函数模型 (11.5.1) 写成如下形式

$$a_t = Y_t - \sum_{j=1}^{\infty} P_j Y_{t-j} - \sum_{j=1}^{\infty} Q_j X_{t-j} \qquad (b > 0) \tag{11.5.9}$$

(应该注意，如果领先指标 X 和输出 Y 之间的传递函数对于 $j < b$ 有 $\nu_j = 0$, 那么，在 (11.5.9) 中的 $Q_1, Q_2, \cdots, Q_{b-1}$ 都是零。)

现在 (11.5.1) 可以写成

$$a_t = (1 - \sum_{j=1}^{\infty} P_j B^j) Y_t - \sum_{j=1}^{\infty} Q_j B^j X_t$$

与 (11.5.1) 式对比，表明权数 P 和 Q 可以通过在如下表达式中

$$\theta(B)(1 - \sum_{j=1}^{\infty} P_j B^j) = \varphi(B)$$

$$\theta(B)\delta(B) \sum_{j=1}^{\infty} Q_j B^j = \varphi(B)\omega(B)B^b$$

令系数值相等而得到。在 (11.5.9) 中用 $t+l$ 代替 t, 并在原点 t 取条件期望，我们便有如下形式的提前 l 期预报

$$\hat{Y}_t(l)=\sum_{j=1}^{\infty}P_j[Y_{t+l-j}]+\sum_{j=1}^{\infty}Q_j[X_{t+l-j}] \tag{11.5.10}$$

现在，提前 1 期的预报是

$$\hat{Y}_t(1)=\sum_{j=1}^{\infty}P_jY_{t+1-j}+\sum_{j=1}^{\infty}Q_jX_{t+1-j}$$

同样，在 (11.5.10) 式中方括号里的量或者是序列 X 和 Y 的已知值，或者是这些已知值的线性函数所作的预报值。

因此我们可以把预报值写成用时刻 t 已经发生的序列值来表示的形式:

$$\hat{Y}_t(l)=\sum_{j=1}^{\infty}P_j^{(l)}Y_{t+1-j}+\sum_{j=1}^{\infty}Q_j^{(l)}X_{t+1-j} \tag{11.5.11}$$

其中的系数 $P_j^{(l)},Q_j^{(l)}$ 可按如下方式递推求得:

$$\begin{aligned}
&P_j^{(1)}=P_j \quad , \quad Q_j^{(1)}=Q_j\\
&P_j^{(l)}=P_{j+l-1}+\sum_{h=1}^{l-1}P_hP_j^{(l-h)}\\
&Q_j^{(l)}=Q_{j+l-1}+\sum_{h=1}^{l-1}\{P_hQ_j^{(l-h)}+Q_h\pi_j^{(l-h)}\}
\end{aligned} \tag{11.5.12}$$

11.5.2 煤气炉输出 CO_2 的预测

为了说明起见，考虑图 11.1 所示煤气炉数据，对于这个例子，拟合模型 (见节 1.4.1) 是

$$Y_t(l)=\frac{-(0.53+0.37B+0.51B^2)}{1-0.57B}X_{t-3}+\frac{a_t}{1-1.53B+0.63B^2}$$

表 11.10　煤气炉模型的权数 π, P 和 Q

j	π_j	P_j	Q_j	j	π_j	P_j	Q_j
1	1.97	1.53	0	7	0	0	−0.07
2	−1.37	−0.63	0	8	0	0	−0.04
3	0.34	0	−0.53	9	0	0	−0.02
4	0	0	0.14	10	0	0	−0.01
5	0	0	−0.20	11	0	0	−0.01
6	0	0	0.43				

及 $(1-1.97B+1.37B^2-0.34B^3)X_t=\alpha_t$ 。写成 (11.5.4) 形式的预报函数为

$$\begin{aligned}\hat{Y}_t=[Y_{t+l}]=&2.1[Y_{t+l-1}]-1.5021[Y_{t+l-2}]+0.3591[Y_{t+l-3}]\\&-0.53[X_{t+l-3}]+0.4409[X_{t+l-4}]-0.2778[X_{t+l-5}]\\&+0.5472[X_{t+l-6}]-0.3213[X_{t+l-7}]\\&+[a_{t+l}]-0.57[a_{t+l-1}]\end{aligned}$$

图 11.9 给出了在原点 $t=206$ 所作提前期 $l=1,2,\cdots,12$ 的预测值。该模型的权数 π, P 和 Q 在表 11.10 中给出。

图 11.9 还给出提前 5 期的预报权数 $P_j^{(5)}$ 和 $Q_j^{(5)}$ 。(11.5.3) 的权数 ν_i 和 ψ_i 在表 11.11 中列出。由分别在节 11.2.2 和 11.4.7 中得到的估计值 $\hat{\sigma}_\alpha^2=0.0353$ 和 $\hat{\sigma}_a^2=0.0561$, 可以用 (11.5.6) 得出预报误差方差及在图 11.9 中所示的 50% 和 95% 的概率限。

为了说明在预报中采用领先指标的好处, 假设只有序列 Y_t 是可以得到的, 对这个序列采用通常的识别和拟合过程, 表明它

表 11.11　煤气炉模型的权数 ν 和 ψ

i	ν_i	ψ_i	i	ν_i	ψ_i
0	0	1	6	−5.33	0.89
1	0	1.53	7	−6.51	0.62
2	0	1.71	8	−6.89	0.39
3	−0.53	1.65	9	−6.57	0.20
4	−1.72	1.45	10	−5.77	0.06
5	−3.55	1.18	11	−4.73	−0.03

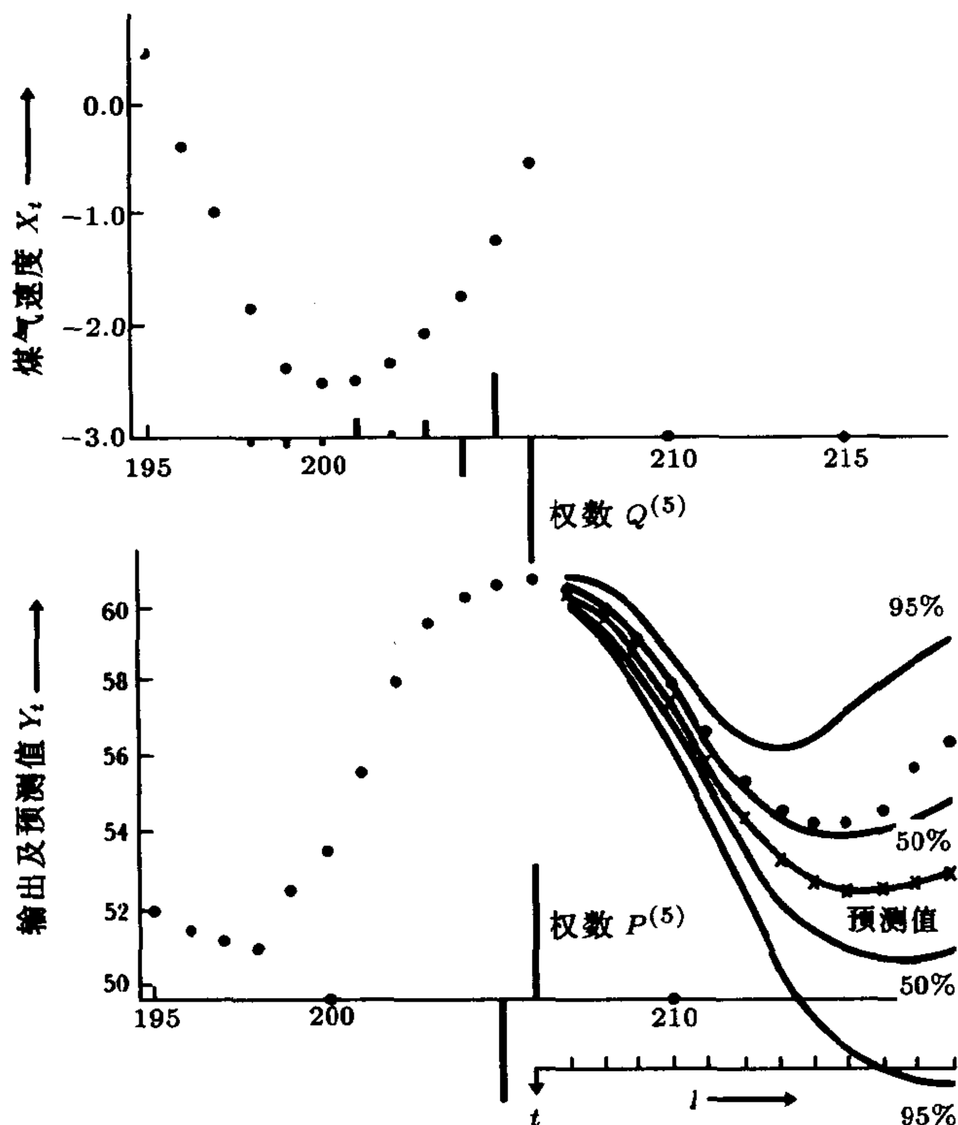


图 11.9 使用输出和输入序列对煤气炉 CO_2 输出的预测

可以由下面的 ARMA(4,2) 过程较好地描述:

$$(1-2.42B+2.38B^2-1.16B^3+0.23B^4)Y_t=(1-0.31B+0.47B^2)\epsilon_t$$

这里 $\hat{\sigma}_\epsilon^2=0.1081$ 。表 11.12 给出了使用和不使用领先指标时预报误差的估计标准差. 正如所料, 对于较短的领先时间, 领先指标的使用能够得到高精度的预报。

单变量建模的检验　为了进一步对于时间序列 Y_t 单变量建模的效果, 我们可以利用附录 A4.3 中的结果, 得出的 Y_t 的单变量 ARIMA 模型, 该模型隐含在 Y_t 和 X_t 之间传递函数模型以及

表 11.12 使用和不使用领先指标时对预报误差估计的标准差

l	有领先指标	无领先指标	l	有领先指标	无领先指标
1	0.23	0.33	7	1.52	2.74
2	0.43	0.77	8	1.96	2.86
3	0.59	1.30	9	2.35	2.95
4	0.72	1.82	10	2.65	3.01
5	0.86	2.24	11	2.87	3.05
6	1.12	2.54	12	3.00	3.08

X_t 的单变量 AR(3) 模型中，这些模型是

$$\begin{aligned}(1-0.57B)&(1-1.53B+0.63B^2)Y_t\\ &=-(0.53+0.37B+0.51B^2)(1-1.53B+0.63B^2)X_{t-3}\\ &\quad+(1-0.57B)a_t \end{aligned} \tag{11.5.13}$$

但是，由于在 X_t 的 AR(3) 模型中

$$\varphi_x(B)=1-1.97B+1.37B^2-0.34B^2 \simeq (1-1.46B+0.60B^2)(1-0.52B)$$

则 (11.5.13) 的右端近似简化为

$$(-0.53+0.37B+0.51B^2)(1-0.52B)^{-1}\alpha_{t-3}+(1-0.57B)a_t$$

从而我们得到

$$\begin{aligned}(1-0.52B)&(1-0.57B)(1-1.53B+0.63B^2)Y_t\\ &=-(0.53+0.37B+0.51B^2)\alpha_{t-3}+(1-0.52B)(1-0.57B)a_t\end{aligned}$$

附录 A4.3 的结果表明，最后这个等式的右端含有一个表示为 $(1-\theta_1 B-\theta_2 B^2)\epsilon_t$ 的 MA(2) 模型，它具有非零的自协方差 $\gamma_0=0.1516$，$\gamma_1=-0.0657$，及 $\gamma_2=0.0262$。因此可知，隐含的 Y_t 的单变量模型是 ARMA(4,2)，近似的 AR 算子为 $(1-2.62B+2.59B^2-1.14B^3+0.19B^4)$，由附录 A6.2 的方法可知， MA(2) 算子是 $(1-0.44B+0.21B^2)$，且有 $\sigma_\epsilon^2=0.1220$。该模型的结果与实际上对序列 Y_t 识

别和拟合出的单变量模型非常一致，这样就对煤气炉数据所确定出的传递函数－噪声模型给出了补充的检验方法，同时也提供了进一步的支持.

11.5.3 使用领先指标对非平稳销售额数据的预报

作为第二个说明示例，我们考虑销售额 Y_t 数据及有关的领先指标 X_t，这些数据绘在图 11.10 中，并在第五部分“时间序列汇集”中作为序列 M 给出. 这类数据在商业预测中是很典型的，可以用下面的非平稳模型[①]很好地拟合：

$$y_t = 0.035 + \frac{4.82x_{t-3}}{1-0.72B} + (1-0.54B)a_t$$
$$x_t = (1-0.32B)\alpha_t$$

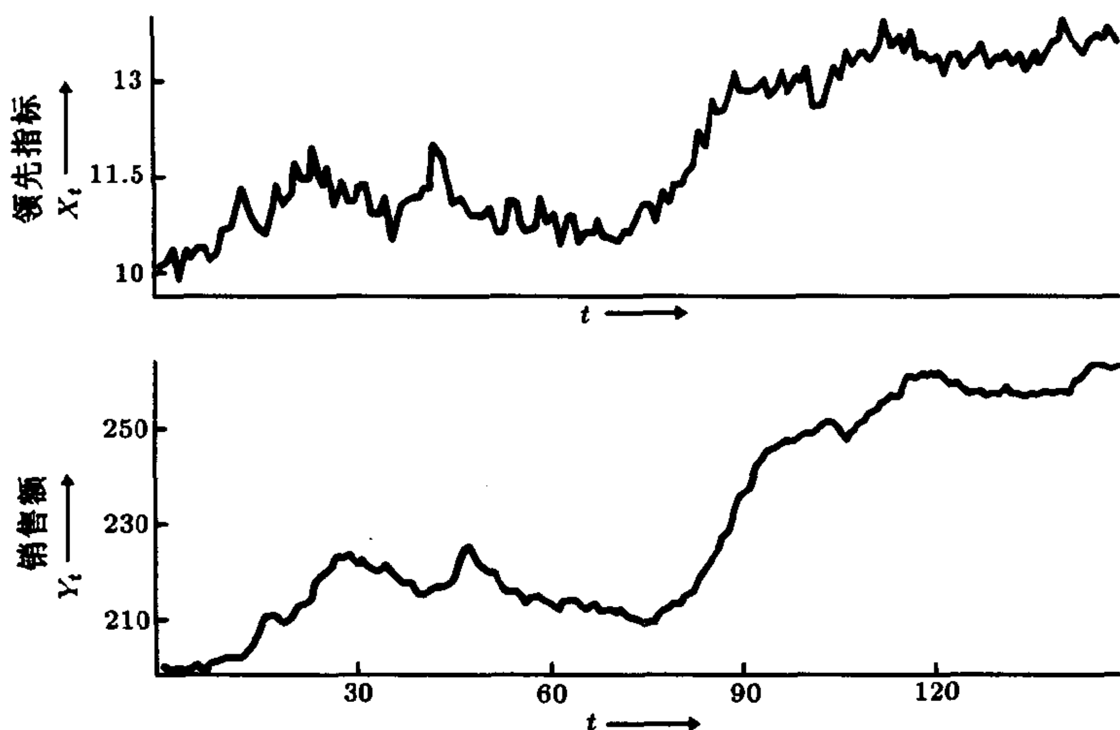


图 11.10 具有领先指标的销售额数据

---

① 所用数据的后一部分即为所列出的序列 M.

这里， y_t 和 x_t 是序列的一阶差分。形如 (11.5.4) 的预报函数是

$$\hat{Y}_t(l)=[Y_{t+l}]=1.72[Y_{t+l-1}]-0.72[Y_{t+l-2}]+0.0098+4.82[X_{t+l-3}]$$
$$-4.82[X_{t+l-4}]+[a_{t+l}]-1.26[a_{t+l-1}]+0.3888[a_{t+l-2}]$$

图 11.11 给出在时间原点 $t=89$ 作出领先时间 $l=1,2,\cdots,12$ 的预报。权数 ν_j 和 ψ_j 在表 11.13 给出。利用由拟合上述模型得到的估计量 $\hat{\sigma}_\alpha^2=0.0676$ 和 $\sigma_a^2=0.0484$, 可从 (11.5.6) 求出预报误差的方差。 50% 和 95% 的概率限见图 11.11. 在这个特定的例子中我们看到，利用领先指标可以得到领先时间为 $l=1,2$ 和 3 的十分精确的预报。该模型的权 π,P 和 Q 在表 11.4 中给出。领先

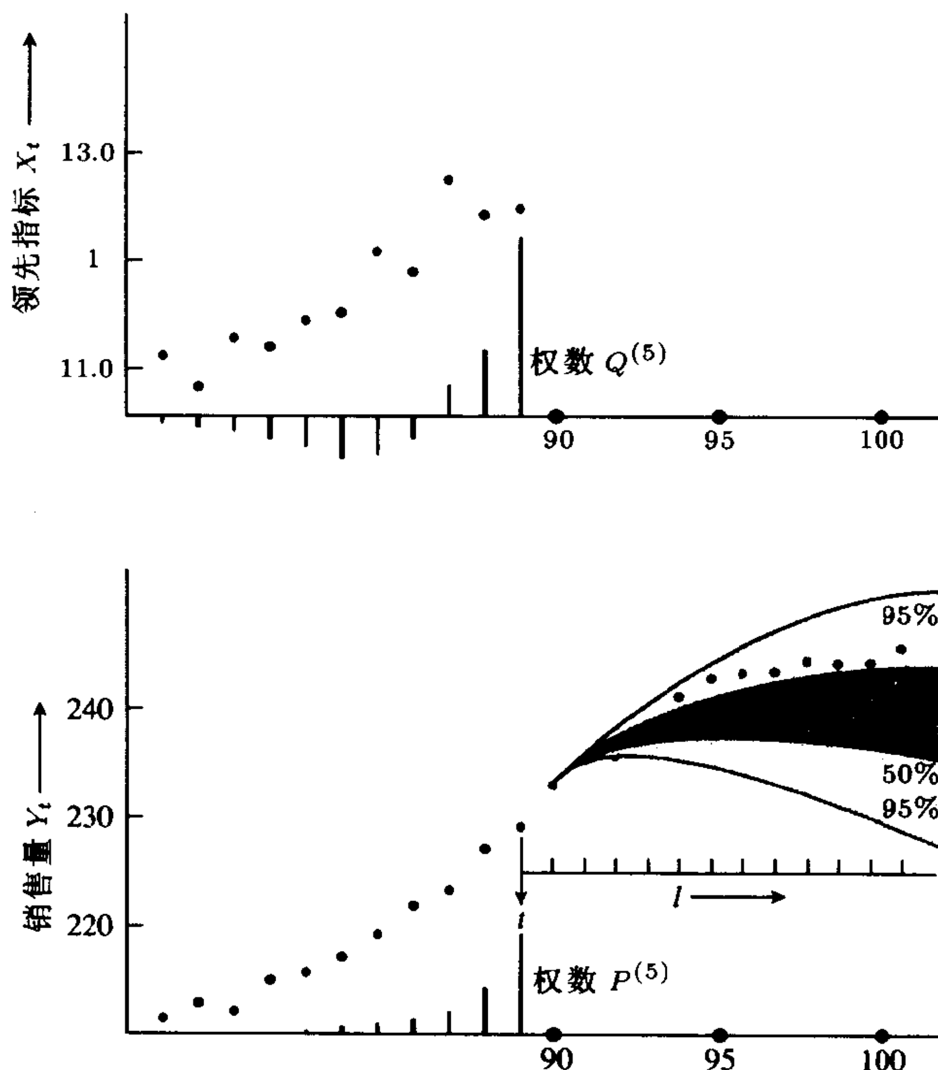


图 11.11 销售量在原点 t=89 的预报值及领先 5 步预报的权数 P 和 Q

表 11.13　非线性模型的权数 ν 和 ψ

j	ν_j	ψ_j	j	ν_j	ψ_j
0	0	1	6	9.14	0.46
1	0	0.46	7	9.86	0.46
2	0	0.46	8	10.37	0.46
3	4.82	0.46	9	10.75	0.46
4	6.75	0.46	10	11.02	0.46
5	8.14	0.46	11	11.21	0.46

表 11.14　非线性模型的权数 π,P 和 Q

j	π_j	P_j	Q_j	j	π_j	P_j	Q_j
1	0.68	0.46	0	9	0.00	0.00	−0.74
2	0.22	0.25	0	10	0.00	0.00	−0.59
3	0.07	0.13	4.82	11	0.00	0.00	−0.29
4	0.02	0.07	1.25	12	0.00	0.00	−0.13
5	0.01	0.04	−0.29	13	0.00	0.00	−0.06
6	0.00	0.02	−0.86	14	0.00	0.00	−0.02
7	0.00	0.01	−0.97	15	0.00	0.00	0.00
8	0.00	0.01	−0.89				

提前 5 期预报值相应的权数 $P_j^{(5)}$ 和 $Q_j^{(5)}$ 见图 11.11。

11.6 估计传递函数有关的试验设计方面的问题

在一些工程应用中，我们可以精心地选择输入 X_t 的形式，以便在传递函数 - 噪声模型

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t$$

中获得参数的良好估计。传递函数的估计等价于“动态回归”模型的估计，可使用的方法也与普通非动态回归非常相似，因此可以预料，二者将面临着同样的问题。[34]

和静态回归一样，弄清楚研究的目的是至关重要的。在一些情形下，我们要回答的问题是：如果输入 X 只能被观测 (而不能加以干预)，在过程运行的正常条件下，有关输出 Y 现在和未来的

特性我们能够知道些什么？在其它场合，适当的问题是：若 X 按某种特定的方式变化，这会导致输出 Y 现在和未来的特性会有怎样的变化？为了回答这两个问题，我们所需要数据的类型是不一样的．为了明确回答第一个问题，我们必须使用对系统正常运行进行观测而不干预所得来的数据．以此相反，为了明确回答第二个问题，对系统输入精心设计的变化应反映在数据中，即数据必须由经过设计的试验来特别地生成。

显然，如果 X 是作为控制变量，即可用来控制输出的变量，那么我们就需要回答第二个问题。为了得到一个因果关系参数的有效估计，我们要弄清楚应该怎样去设计试验。考察一下分析问题的假定是必要的。

一个重要的假定是 X 和 N 的分布相互独立，当这一假定被破坏时：

1. 我们得到的估计一般不是一致的．具体来说，就是当样本容量很大时，估计值并不收敛于真值，而是收敛于和真值相差某个未知量的其它值。

2. 这一假定的违反并不能由考察数据而探知，在任何具体的场合，独立性假定度有可能不满足，这是很麻烦的．唯一能够保证这一假定成立的办法是：精心设计试验，而不是使用简单“发生”的数据．具体来说，我们必须精心产生一个输入 X_t，并将其馈入到过程中去。我们知道 X_t 是与 N_t 无关的，因为 X_t 是通过某个外部随机过程来生成的。

输入 X_t 当然可以是自相关的，唯一必要的是它不应与 N_t 互相关。为了满足这一要求，例如，我们可以抽取一组随机偏差 α_t，用以生成任何预期的随机过程 $X_t = \psi_X(B)\alpha_t$ 。

另外，我们还可以选择一种“固定”的设计，例如在 11.4.2 节中使用的因子设计，并使其每轮运行的顺序随机化。在附录 A11.2 中包含了一些基础设计问题的初步讨论，足以解决在实际中选择“最优”随机输入的一些困难。特别地，是在更广的意义下存在

下面的困难：(1) 对于"最优性"确定出一个切实的标准是困难的，(2)"最优"输入的选择依赖于那些需要进行最优估计的未知参数值. 通常，白噪声输入对于简化识别具有明显的好处，如果对于所研究的系统任何明确的信息都不知道，那么，白噪声为输入提供了一种可行的初始选择。

A 11.1 互谱分析用于传递函数模型的识别

在本附录中，我们给出检验传递函数模型的另一种方法，该方法可以建立在谱分析的基础上，不要求对输入的白噪化，且很容易推广到多维输入的情形。

A11.1.1 单输入传递函数模型的识别

假设传递函数 $v(B)$ 被定义为：对于非负整数 j, 允许有非零的脉冲响应函数 v_j, 即

$$v(B) = \sum_{k=-\infty}^{\infty} v_k B^k$$

那么，相应于 (11.2.3), 传递函数 - 噪声模型为

$$y_t = \theta(B) x_t + n_t$$

等式 (11.2.5) 成为

$$\gamma_{xy}(k) = \sum_{j=-\infty}^{\infty} v_j \gamma_{xx}(k-j) \qquad (k = 0, \pm 1, \pm 2, \cdots) \qquad \text{(A11.1.1)}$$

我们现在来定义一个互协方差生成函数

$$\gamma^{xy}(B) = \sum_{-\infty}^{\infty} \gamma_{xy}(k) B^k \qquad \text{(A11.1.2)}$$

它类似于自协方差生成函数 (3.1.10)．用 B^k 去乘 (A11.1.1) 并求和，我们得到

$$\gamma^{xy}(B) = v(B)\gamma^{xx}(B) \qquad \text{(A11.1.3)}$$

如果我们现在将 $B=e^{j2\pi f}$ 代入到 (A11.1.2) 中，便得到输入与输出之间的互谱 $p_{xy}(f)$ 。在 (A11.1.3) 中做同样的代换，得到

$$v(e^{-i2\pi f})=\frac{p_{xy}(f)}{p_{xx}(f)} \qquad (-\frac{1}{2}<f<\frac{1}{2}) \tag{A11.1.4}$$

其中

$$v(e^{-i2\pi f})=G(f)e^{i2\pi\phi(f)}=\sum_{-\infty}^{\infty}v_k e^{-i2\pi fk} \tag{A11.1.5}$$

称为系统的频率响应函数，它是脉冲响应函数的富里叶变换。由于 $v(e^{-i2\pi f})$ 是复值的，我们将其写为增益函数 $G(f)$ 与相位函数 $\phi(f)$ 的乘积。 (A11.1.4) 式表明，频率响应函数是互谱与输入谱之比。估计频率响应函数 $v(e^{-i2\pi f})$ 的方法在 [122] 中有述。知道了 $v(e^{-i2\pi f})$, 则脉冲响应函数 v_k 可由下式得出

$$v_k=\int_{-1/2}^{1/2}v(e^{-i2\pi f})e^{i2\pi fk}df \tag{A11.1.6}$$

利用类似的方法，噪声 n_t 的自协方差生成函数为

$$\gamma^{nn}(B)=\gamma^{yy}(B)-\frac{\gamma^{xy}(B)\gamma^{xy}(F)}{\gamma^{xx}(B)} \tag{A11.1.7}$$

将 $B=e^{-i2\pi f}$ 代入 (A11.1.7), 我们得到噪声谱的表达式

$$p_{nn}(f)=p_{yy}(f)[1-\kappa_{xy}^2(f)] \tag{A11.1.8}$$

这里,

$$\kappa_{xy}^2(f)=\frac{|p_{xy}(f)|^2}{p_{xx}(f)p_{yy}(f)}$$

$\kappa_{xy}(f)$ 是凝聚谱，其特性就象在每个频率 f 处的相关系数。知道了噪声谱，则噪声的自协方差函数 $\gamma_{nn}(k)$ 可由下式得到

$$\gamma_{nn}(k)=2\int_0^{1/2}p_{nn}(f)\cos(2\pi fk)df$$

将 [122] 所述的谱估计值代进去，就可以得到脉冲响应权数 v_k 及噪声自相关函数的估计值。如节 11.2.1 和 6.2.1 中所述，它们可用于识别传递函数模型和噪声模型。

A11.1.2 多输入传递函数模型的识别

我们现在对模型

$$\begin{aligned} Y_t &= v(B)X_{t-b} + N_t \\ &= \delta^{-1}(B)\omega(B)X_{t-b} + N_t \end{aligned}$$

加以推广，使之允许有几个输入 $X_{1,t}, X_{2,t}, \cdots, X_{m,t}$ 。于是

$$Y_t = v_1(B)X_{1,t} + \cdots + v_m(B)X_{m,t} + N_t \tag{A11.1.9}$$

$$= \delta_1^{-1}(B)\omega_1(B)X_{1,t-b_1} + \cdots + \delta_m^{-1}(B)\omega_m(B)X_{m,t-b_m} + N_t \tag{A11.1.10}$$

这里 $v_j(B)$ 是将 $X_{j,t}$ 与输出联系起来的脉冲响应权数的生成函数。与前面一样，我们假设经差分后 (A11.1.9) 可以写成

$$y_t = v_1(B)x_{1,t} + \cdots + v_m(B)x_{m,t} + n_t$$

用 $x_{1,t-k}, x_{2,t-k}, \cdots, x_{m,t-k}$ 依次来乘上式，并取期望以构成生成函数，我们得到

$$\begin{aligned} \gamma^{x_1y}(B) &= v_1(B)\gamma^{x_1x_1}(B) + v_2(B)\gamma^{x_1x_2}(B) + \cdots + v_m(B)\gamma^{x_1x_m}(B) \\ \gamma^{x_2y}(B) &= v_1(B)\gamma^{x_2x_1}(B) + v_2(B)\gamma^{x_2x_2}(B) + \cdots + v_m(B)\gamma^{x_2x_m}(B) \\ &\vdots \\ \gamma^{x_my}(B) &= v_1(B)\gamma^{x_mx_1}(B) + v_2(B)\gamma^{x_mx_2}(B) + \cdots + v_m(B)\gamma^{x_mx_m}(B) \end{aligned} \tag{A11.1.11}$$

将 $B = e^{-i2\pi f}$ 代入可得到谱方程。例如 $m = 2$ 时有

$$\begin{aligned} p_{x_1y}(f) &= H_1(f)p_{x_1x_1}(f) + H_2(f)p_{x_1x_2}(f) \\ p_{x_2y}(f) &= H_1(f)p_{x_2x_1}(f) + H_2(f)p_{x_2x_2}(f) \end{aligned}$$

频率响应函数 $H_1(f)=v_1(e^{-i2\pi f}), H_2(f)=v_2(e^{-i2\pi f})$ 可依 [122] 所述计算。因此，脉冲响应函数可用其逆变换 (A11.1.6) 来得到。

A 11.2 选择输入以得到最优的参数估计

假设可以使动态系统的输入服从于一个选定的外加随机过程，例如，可以是自回归过程，滑动平均过程或白噪声。为了说明这个随机过程最优选择所涉及的问题，我们只须考虑一个基本的例子。

A11.2.1 对一个简单系统最优输入的设计

假设一个正在研究的系统，其传递函数模型假设为

$$Y_t=\beta_1Y_{t-1}+\beta_2X_{t-1}+a_t \qquad (|\beta_1|<1) \tag{A11.2.1}$$

其中 a_t 是白噪声。还假设输入和输出过程是平稳的， X_t 和 Y_t 记这些过程对其相应均值的偏差。在大样本情形，并与任意概率限相联系， β_1 和 β_2Bayes–HPD 区域的近似面积，以及相应的置信区域都与 $\Delta^{-1/2}$ 成比例，这里

$$\Delta=\begin{vmatrix} E[Y_t^2] & E[Y_tX_t] \\ E[Y_tX_t] & E[X_t^2] \end{vmatrix}$$

我们将着手寻求使得上述区域的面积为最小且 Δ 为最大的设计。现在

$$\begin{aligned} E[Y_t^2]&=\sigma_Y^2=\sigma_X^2\beta_2^2\frac{1+2q}{1-\beta_1^2}+\frac{\sigma_a^2}{1-\beta_1^2} \\ E[Y_tX_t]&=\sigma_X^2\frac{\beta_2}{\beta_1}q \\ E[X_t^2]&=\sigma_X^2 \end{aligned} \tag{A11.2.2}$$

这里

$$q=\sum_{i=1}^{\infty}\beta_1^i\rho_i\ ,\qquad \sigma_X^2\rho_i=E[X_tX_{t-i}]$$

行列式的值用 σ_X^2 来表示可写成

$$\Delta=\frac{\sigma_X^2\sigma_a^2}{1-\beta_1^2}+\frac{\beta_2^2\sigma_X^4}{(1-\beta_1^2)^2}-\frac{\sigma_X^4\beta_2^2}{\beta_1^2}\left(q-\frac{\beta_1^2}{1-\beta_1^2}\right)^2 \qquad \text{(A11.2.3)}$$

因此，可以预料，通过使 σ_X^2 越大，可以使区域的面积越小 (即使得输入变量在一个较大的范围变化)．在实际中， X 允许的变化量级可能是有限的。我们先假设 σ_X^2 保持在某一确定值，由此着手讨论。

σ_X^2 固定时的求解　当 $(1-\beta^2)>0$ 时，对任意 σ_X^2, 我们由 (A11.2.3) 看到，通过令

$$q=\frac{\beta_1^2}{1-\beta_1^2}$$

可以使 Δ 极大化，即

$$\beta_1\rho_1+\beta_1^2\rho_2+\beta_1^3\rho_3+\cdots=\beta_1^2+\beta_1^4+\beta_1^6+\cdots$$

对于给定的 β_1, 有无穷多种方法可使该等式成立，其中的一个解是：

$$\rho_i=\beta_1^i$$

因此，对于固定的 σ_X^2, 使 Δ 极大化的一种办法就是强使输入服从下面的自回归过程

$$(1-\beta_1B)X_t=\epsilon_t$$

这里， ϵ_t 是具有方差 $\sigma_\epsilon^2=\sigma_X^2(1-\beta_1^2)$ 的白噪声过程。

σ_Y^2 **固定时的求解** 截至目前为止我们假设 σ_Y^2 是不受约束的。在某些场合，我们可能会希望在输出（而不是输入）中避免太大的变动。假设 σ_Y^2 维持在某个可接受的固定值，而 σ_X^2 不受约束。那么，行列式 Δ 可以用 σ_Y^2 表示为

$$\Delta + \frac{\sigma_Y^4}{\beta_2^2}\left[\frac{\sigma_Y^2 - \sigma_a^2}{\sigma_Y^2} - \frac{\beta_1^2}{s^2}\left(\frac{q+s}{1+2q}\right)^2\right] \tag{A11.2.4}$$

这里

$$s = \frac{\beta_1^2 r}{1+\beta_1^2 r} \tag{A11.2.5}$$

及

$$r = \frac{\sigma_Y^2}{\sigma_Y^2 - \sigma_a^2} \tag{A11.2.5}$$

为达到极大，可令

$$q = -s = -\beta_1^2 r/(1+\beta_1^2 r) \tag{A11.2.7}$$

即

$$\beta_1\rho_1 + \beta_1^2\rho_2 + \beta_1^3\rho_3 + \cdots = -\beta_1^2 r + \beta_1^4 r^2 - \beta_1^6 r^3 + \cdots$$

这个等式也有无穷多种方式来满足。我们特别令

$$\rho_i = (-\beta_1 r)^i \tag{A11.2.8}$$

并强使输入服从自回归过程

$$(1+\beta_1 rB)X_t = \epsilon_t \tag{A11.2.9}$$

这里 ϵ_t 是具有方差 $\sigma_\epsilon^2 = \sigma_X^2(1-\beta_1^2 r^2)$ 的白噪声过程。由于 r 基本上是正的，所以该自回归过程中参数 $\beta_1 r$ 的符号与 σ_X^2 固定时为最优输出而得出的符号相反。

$\sigma_Y^2 \times \sigma_X^2$ 固定时的求解　在实际中可能会遇到这样的情况：输入和输出都应避免过大的变化．这时， X 的方差降低某个固定的百分比将等价于要求 Y 的方差降低同样的百分比，因此，在乘积 $\sigma_X^2 \times \sigma_Y^2$ 为某一固定值的条件下极大化 Δ 是可行的．行列式成为

$$\Delta = \sigma_X^2 \sigma_Y^2 - \frac{\sigma_X^4 \beta_2^2 q^2}{\beta_1^2} \qquad \text{(A11.2.10)}$$

对于固定的 $\sigma_X^2 \sigma_Y^2$，该式仅在 $q = 0$ 时极大化。我们再次看到，存在着无穷多个解。但是，采用白噪声输入，不论 β_1 的值是什么都可以使 Δ 达到极大。对于这样的输入，利用 (A11.2.2),σ_X^2 应是下面方程的正根

$$\sigma_X^4 \beta_2^2 + \sigma_X^2 \sigma_a^2 - k(1 - \beta_1^2) = 0 \qquad \text{(A11.2.11)}$$

其中 $k = \sigma_X^2 \sigma_Y^2$ 。

A11.2.2 数值例子

假设我们所研究的一阶动态系统 (A.11.2.1) 有 $\beta_1 = 0.50, \beta_2 = 1.00$, 从而

$$Y_t = 0.50Y_{t-1} + 1.00X_{t-1} + a_t$$

这里 $\sigma_a^2 = 0.2$ 。

σ_X^2 固定，σ_Y^2 不受约束　首先假设在 $\sigma_X^2 = 1.0$ 时来选择使 Δ 极大化的设计。那么，输入 X_t 的一个最优选择是自回归过程

$$(1 - 0.5B)X_t = \epsilon_t$$

其中白噪声过程 ϵ_t 具有方差 $\sigma_\epsilon^2 = \sigma_X^2(1 - \beta_1^2) = 0.75$ 由 (A11.2.2), 输出的方差 σ_Y^2 是 2.47, 这样的方案可得到关于 β_1 和 β_2 的 Bayes 区域，其面积与 $\Delta^{1/2} = 0.70$ 成正比。

σ_Y^2 固定，σ_X^2 不受约束　上面的方案在假设输入方差为 $\sigma_X^2 = 1$ 而输出方差不受约束时是最优的．结果是输出的方差为

$\sigma_Y^2 = 2.49$ 。与此相反，如果假设输入方差不受约束，而具有固定的输出方差 $\sigma_Y^2 = 2.49$，我们当然可以做得好得多。事实上，利用 (A11.3.6) 有 $r = 1.087$，因此，由 (A11.2.9)，对于这个无约束输入的一个最优选择是自回归过程

$$(1 + 0.5B)X_t = \epsilon_t$$

这时，ϵ_t 是白噪声过程，其方差 $\sigma_\epsilon^2 = 2.05$ 。现在输入的方差 σ_X^2 增加到了 2.91，而度量 Bayes 区域面积的量 $\Delta^{-1/2}$ 将减少到 $\Delta^{-1/2} = 0.42$ 。

乘积 $\sigma_Y^2 \times \sigma_X^2$ 固定　最后，我们来考虑一个试图同时控制 σ_Y^2 和 σ_X^2 的方案，即通过固定 $\sigma_Y^2 \times \sigma_X^2$ 以使 Δ 极小化。在前面 σ_Y^2 固定的例子中，我们得出 $\Delta^{-1/2} = 0.42$ ，且有 $\sigma_X^2 = 2.91$ 及 $\sigma_Y^2 = 2.49$，因此相应乘积为 $2.91 \times 2.49 = 7.25$ 。如果我们的目标是在保持该乘积等于 7.25 时使 $\Delta^{-1/2}$ 极小，那么，我们可以不用 β_1 的信息而选一个白噪声输入 $X_t = \epsilon_t$ 作为最优选择。由 (A11.2.11)，$\sigma_X^2 = \sigma_\epsilon^2 = 2.29, \sigma_Y^2 = 3.16$，正如所料，在这种情形下 $\Delta^{-1/2} = 0.37$ 比前面的例子中稍小。

这个例子值得用谱的观点来考虑。为了在 σ_X^2 固定时实现最优，我们使用了 ϕ 为正的自回归输入，它在低频部分有较高的能量。由于系统的增益在低频部分是较高的，这就实现了从 X 到 Y 的最大传递，从而导致了 Y 的较大变差。当 σ_Y^2 固定时，我们引入了 ϕ 为负的自回归过程作为输入，它在高频部分具有较高的能量。由于在高频部分存在从 X 到 Y 的极小化传递，因此，在 X 中的干扰必然在这些频率上被放大。当乘积 $\sigma_X^2 \times \sigma_Y^2$ 固定时，给出了"折衷"的白噪声输入，且不需要 β_1 的信息。最终的极大化等价于估计值 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 之间相关的极小化。事实上，当采用白噪声输入时，这两个估计值之间的相关为零。

结论　以上研究表明：

1. 设计的最优选择在很大程度上取决于我们如何给“最优”下定义。

2. 在固定 σ_X^2 和固定 σ_Y^2 的两种情形下，最优选择要求特定的随机过程，其参数是未知动态参数的函数。因此，我们处于通常所说的自相矛盾的境地，我们有可能通过对数据的推断做出较好的工作，但这不过是我们想要寻求的答案中已经知道了的一些东西。当我们发现了更多有关参数的信息时，就去改进设计，这种序贯的方法可能是值得进一步研究的。特别地，用一个可能并非最优的输入，比如白噪声，来进行试探性研究，由此而生成的数据可以得到动态参数的初步估计，然后，采用我们前文中的某个准则，利用这些估计值再确定出进一步的输入。

3. 在我们所研究的简单例子中，对于一个可行的优化准则来说，白噪声的使用被证明是最好的，若将白噪声作为输入时，不要求参数的先验信息。

12

干预分析模型和异常值检测

时间序列常常会受到特殊事件及态势的的影响，诸如政治变更，罢工，广告促销，环境法规等等，我们把类似的这些事件称为干预事件。在本章中我们叙述干预分析的方法，并解释这些干预的影响。为此，要用到前一章的传递函数模型，不过，在干预分析模型中，输入序列是以简单的脉冲或阶跃示性函数来表示事件是否发生。一开始，假设这些干预事件发生的时间是已知的，稍后，对于时间序列在某未知时点上异常或非正常行为的检测方法也进行了讨论。

12.1 干预分析方法

12.1.1 干预分析的模型

在进行干预分析时，要假设干预事件发生于时间序列的某个已知时点。使我们感兴趣的是：与该事件相联系的被研究序列 Y_t 中是否发生明显的变化，是否有预期的影响。我们考虑用传递函数模型来模拟干预影响的特征，并估计干预影响的量级，从而解释在时间序列中与事件有关、可能出现的非正常表现。基于 Box 和 Tiao[59] 的研究，我们所考虑模型的类型具有如下形式

$$Y_t = \frac{\omega(B)B^b}{\delta(B)}\xi_t + N_t \tag{12.1.1}$$

这里，$\mathcal{Y}_t = \delta^{-1}(B)\omega(B)B^b\xi_t$ 代表干预事件的影响，它用确定性输入序列 ξ_t 的形式来表示。N_t 是噪声，它表示在没有干预影响时对序列 Y_t 的观测背景。假设 N_t 服从一个 ARIMA(p,d,q) 模型：$\varphi(B)N_t = \theta(B)a_t$, 其中 $\varphi(B) = \phi(B)(1-B)^d$。第 9 章中给出的乘积季节 ARIMA 模型也可用于表示 N_t, 但是在本章中并没有对季节模型特别注重。

有两类通用的确定性输入变量 ξ_t, 人们发现用它们描述干预事件对时间序列的影响是很有用的。它们都是只取 0 和 1 的示性变量，用以标识干预事件未发生或是发生。其中一类是在时刻 T 的阶跃函数，由下式给出

$$S_t^{(T)} = \begin{cases} 0 & , \quad t < T \\ 1 & , \quad t \geqslant T \end{cases} \tag{12.1.2}$$

这通常用于表示在时刻 T 之后干预的影响仍保留下去的情形。另一类是在 T 时刻的脉冲函数，由下式给出

$$P_t^{(T)} = \begin{cases} 0 & , \quad t \neq T \\ 1 & , \quad t = T \end{cases} \tag{12.1.3}$$

它可以用来表示暂时或瞬时干预的影响，在时刻 T 之后，干预的影响就会消失。在许多场合都要使用这类示性输入变量，这时干预的影响不能表示为定量变量的响应，可能是因为这样的定量变量不存在，或者是要得到该量的度量是不可能、不切实际的。

由于 (12.1.1) 中的示性变量是确定性的，所以与第 11 章中传递函数模型的情形不同，对干预模型算子 $v(B) = \delta^{-1}(B)\omega(B)B^b$ 结构的识别不能以白噪化技术为基础。可行的办法是：通过考虑那些有可能造成变化或影响的机制以及预期变化的隐含形式，由此确定干预模型的形式。另一方面，模型的识别可以通过直接观察数据，以揭示由已知事件造成影响的形式，另外，考察干预项引入之前所拟合模型的残差，有时也能提供补充的证据。

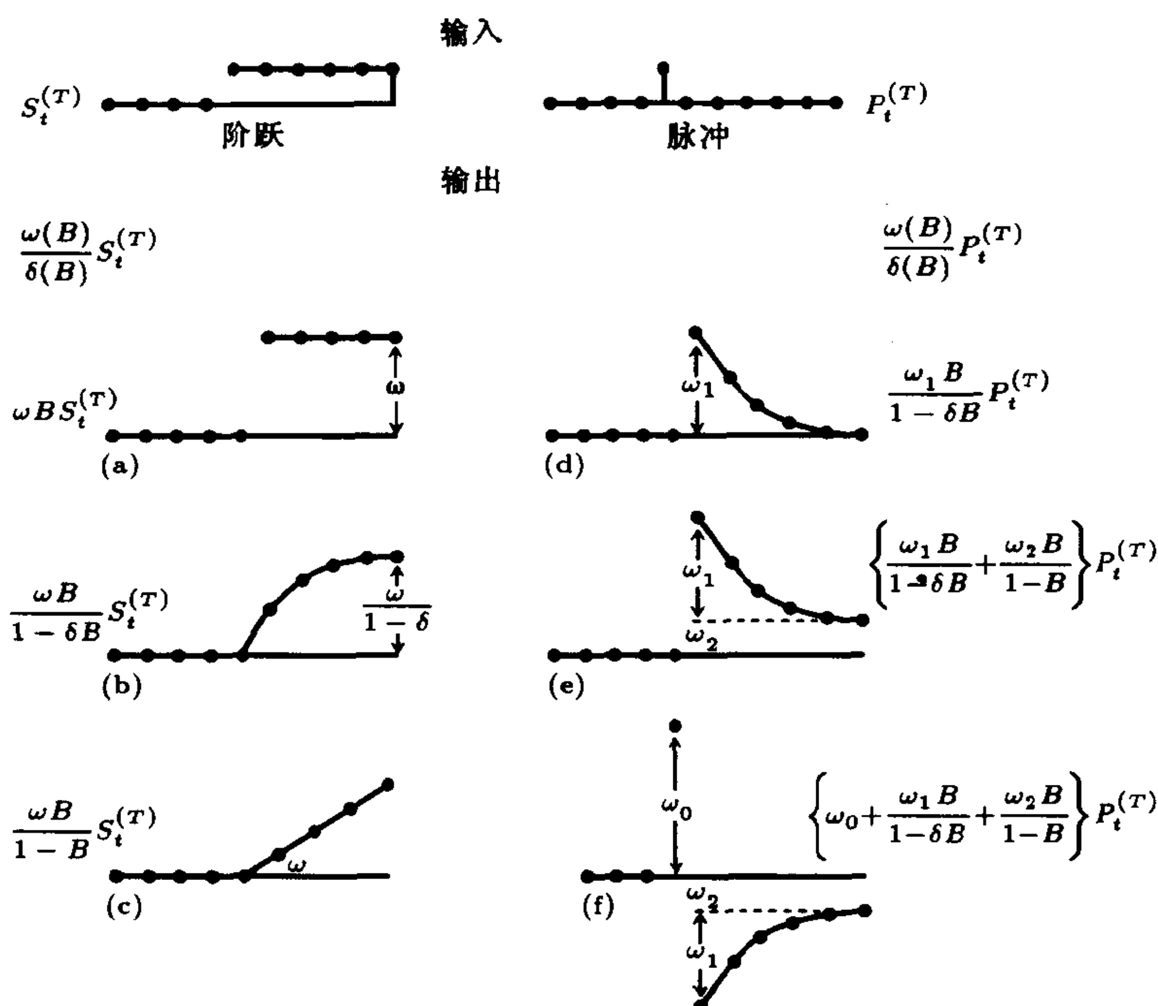


图 12.1 对于阶跃和脉冲输入的响应：(a),(b),(c) 是各种简单传递函数模型对阶阶跃输入的响应；(d),(e),(f) 是一些重要模型对脉冲输入的响应

通过传递函数的不同选择，有若干种不同的响应形式 $\mathcal{Y}_t = \delta^{-1}(B)\omega(B)B^b\xi_t$ 都是可能的。图 12.1 给出了多种简单的传递函数对于阶跃和脉冲示性输入的响应。例如，在图 12.1(a) 中，模型 $\mathcal{Y}_t = \omega BS_t^{(T)}$ 可用于描述在时刻 T 后达到未知水平 ω 的永久阶跃变化，而图 12.1(b) 中的形式

$$\mathcal{Y}_t = \frac{\omega B}{1-\delta B}S_t^{(T)} \qquad (0 < \delta < 1) \tag{12.1.4}$$

意味着 $\mathcal{Y}_t = \omega(1-\delta^{t-T})/(1-\delta), t > T$，它对应于以速度为 δ 的渐近变化，最终将接近水平为 $\omega/(1-\delta)$ 的变化。类似地，图 12.1(d)

中的模型

$$\mathcal{Y}_t = \frac{\omega_1 B}{1-\delta B} P_t^{(T)} \qquad (0<\delta<1) \tag{12.1.5}$$

意味着 $\mathcal{Y}_t = \omega_1 \delta^{t-T-1}, t > T$, 它表示一个突发的“脉冲”式变化，在时刻 T 后达到未知水平 ω_1, 然后，以速度 δ 渐近衰减，回落到干预前的水平因而没有持久的影响。更复杂的响应形式可以通过简单形式的各种线性组合得到，例如，图 12.1(f) 的情形。另外还应注意到，由于 $(1-B)S_t^{(T)} = P_t^{(T)}$, 所以任一包含 $S_t^{(T)}$ 的传递函数模型可以等价地用 $P_t^{(T)}$ 来表示。

关于干预模型还需要注意下面几点。函数 $\mathcal{Y}_t$ 表示的是干预事件对噪声或“背景”序列 N_t 附加的影响。因此，在可能的时候，对噪声模型 $N_t = [\theta(B)/\varphi(B)]a_t$ 可用通常的方法来识别，即对干预时刻之前的时间序列观测值 $Y_t, t<T$ 进行识别，而且在模型 (12.1.1) 中，假设只有序列的水平受干预的影响，N_t 的时间序列模型的形式和参数在干预前后则是相同的。还应该认识到，干预模型参数估计精度可能存在相当大的差别，这取决于噪声 N_t 是否是平稳，以及对于干预影响的假定是持续的还是短暂的。

一般地，对于干预模型

$$Y_t = \frac{\omega(B)B^b}{\delta(B)}\xi_t + \frac{\theta(B)}{\varphi(B)}a_t \tag{12.1.6}$$

其参数估计及相应标准差可以用 11.3 节中传递函数 — 噪声模型的最小二乘估计方法得到。对模型残差的诊断检验与前面讨论的方法类似，也可用于评价所拟合模型的恰当程度。

12.1.2 干预分析的例子

Box 和 Tiao 考虑了 1953 年 7 月到 1972 年 12 月美国消费价格指数 (CPI) 变化率的月度序列。阶段 I 从 1971 年 12 月开始实行了经济控制，历时三个月，其后阶段 II 开始。考虑阶段 I 和阶段 II 由于控制 CPI 的变化率而可能产生的影响是很有意义的。

对于阶段 I 之前的 218 个月度观测值的一阶差分和 CPI 变化率，我们观察其样本自相关函数，联想到噪声模型应有下面形式

$$(1-B)N_t = (1-\theta B)a_t \tag{12.1.7}$$

其极大似然估计为 $\hat{\theta}=0.84$ 及 $\hat{\sigma}_a^2=0.0019$ 。对残差及其自相关的检验表明，该模型并无明显不当。为了着手解决阶段 I 和阶段 II 控制的可能影响这一问题，假设阶段 I 和 II 中 CPI 变化率的水平将引起变动，而噪声模型的形式保持不变。基于这些假设，评价控制效果的恰当模型为

$$Y_t = \omega_1\xi_{1t} + \omega_2\xi_{2t} + \frac{1-\theta B}{1-B}a_t \tag{12.1.8}$$

这里

$$\xi_{1t} = \begin{cases} 1 & , \quad t = 1971 \text{ 年 } 9 \text{ 月，} 10 \text{ 月或 } 11 \text{ 月} \\ 0 & , \quad \text{其它} \end{cases}$$

$$\xi_{2t} = \begin{cases} 1 & , \quad t \geqslant 1971 \text{ 年 } 12 \text{ 月} \\ 0 & , \quad \text{其它} \end{cases}$$

可以得到模型 (12.1.8) 中参数的非线性最小二乘估计及标准差 (括号内)，它们是

$$\hat{\theta} = 0.85(0.05), \hat{\omega}_1 = -0.0022(0.0010), \hat{\omega}_2 = -0.0008(0.0009)$$

因此，分析表明阶段 I 与 CPI 增长率的下降相联系，而阶段 II 的影响则很不确定。

使用干预分析的例子还有许多，其中如：新产汽车发动机设计规定的改变对于洛杉矶地区二氧化碳污染水平的影响 [59]，债务政策的变化对于坏债权数额的影响 [121]，安全带立法对公路死亡率的影响 [31]，阿拉伯石油禁运对美国电力消费的影响 [145] 。

12.1.3 仅有水平变化的简单参数模型的 MLE 性质

对于象 (12.1.8) 这类相对简单的情形，考虑干预参数的极大似然估计的性质，是颇具启发性的。我们考虑简单模型

$$Y_t = \omega\xi_1 + N_t \tag{12.1.9}$$

这里 $N_t = \varphi^{-1}(B)\theta(B)a_t$ 。该模型可以形式地写为

$$\pi(B)Y_t = \omega\pi(B)\xi_t + a_t \tag{12.1.10}$$

其中 $\pi(B) = \theta^{-1}(B)\varphi(B) = 1 - \sum_{i=1}^{\infty}\pi_t B^i$ 。令 $w_t = \pi(B)Y_t$, 及 $x_t = \pi(B)\xi_t$, 我们可以将 (12.1.10) 写成简单线性模型 $w_t = \omega x_t + a_t, t = 1, 2, \cdots, n$ 。所以 ω 的极大似然估计近似为

$$\hat{\omega} = \frac{\sum_{t=1}^{n} x_t w_t}{\sum_{t=1}^{n} x_t^2} \tag{12.1.11}$$

其方差 $\text{var}(\hat{\omega}) = \sigma_a^2 / \sum_{t=1}^{n} x_t^2$ 。

我们来考虑 (12.1.9) 的特殊情形，即 $\xi_t = BS_t^{(T)}$ 是在时刻 T 的阶跃函数。那么，$x_t = \pi(B)BS_t^{(T)} = 1 - \sum_{i=1}^{t-T-1}\pi_i, t > T+1$, 这里 $x_{T+1} = 1$，且 $x_t = 0$, 当 $t < T$ 。为了下面的讨论，我们假设 n 很大，在干预时刻 T 之前或之后都能得到相当多的观测值。

现在假设 (12.1.9) 中的噪声 N_t 服从 IMA(0,1,1) 模型：$(1-B)N_t = (1-\theta B)a_t$, 所以 $\pi(B) = (1-\theta B)^{-1}(1-B)$, 其中 $\pi_i = (1-\theta)\theta^{i-1}, i > 1$ 。于是，$x_t = (1-\theta B)^{-1}(1-B)BS_t^{(T)} = (1-\theta B)^{-1}BP_t^{(T)} = \theta^{t-T-1}, t > T+1$, 从而

$$\sum_{t=1}^{n} x_t^2 = \sum_{t=T+1}^{n} \theta^{2(t-T-1)} = \frac{1-\theta^{2(n-T)}}{1-\theta^2} \simeq \frac{1}{1-\theta^2}$$

另外还有： $w_t=(1-\theta B)^{-1}(1-B)Y_t=Y_t-(1-\theta)\sum_{i=1}^{\infty}\theta^{i-1}-Y_{t-1}=Y_t-\overline{Y}_{t-1}$, 其中 $\overline{Y}_{t-1}$ 是在 t 时刻之前值的指数加权滑动平均。按照 [59] 中的结果，可以证明

$$\sum_{t=1}^{n}x_tw_t=\sum_{t=T+1}^{n}\theta^{t-T-1}w_t\simeq(1+\theta)^{-1}(\sum_{s=0}^{\infty}\theta^sY_{T+1+s}-\sum_{s=0}^{\infty}\theta^sY_{T-s})$$

因此，在这种情形下 ω 的极大似然估计是

$$\hat{\omega}\simeq(1-\theta)(\sum_{s=0}^{\infty}\theta^sY_{T+1+s}-\sum_{s=0}^{\infty}\theta^sY_{T-s}) \tag{12.1.12}$$

方差 $\mathrm{var}[\hat{\omega}]\simeq\sigma_a^2(1-\theta^2)$ 。估计值 $\hat{\omega}$ 可以解释为两个指数加权滑动平均之间的差异，其中一个由干预之后的观测值组成，另一个由干预之前的观测值组成。

现在作为第二种情形，假设噪声不是服从 ARI(1,1,0) 模型，而是 $\pi(B)=(1-\phi B)(1-B)$ 。于是 $x_t=(1-\phi B)(1-B)BS_t^{(T)}=(1-\phi B)BP_t^{(T)}=0,t>T+2$, 这里 $x_{T+1}=1$ 且 $x_{T+2}=-\phi$, 因而 $\sum_{t=1}^{n}x_t^2=1+\phi^2$ 。另外， $w_t=(1-\phi B)(1-B)Y_t$, 得到

$$\begin{aligned}\sum_{t=1}^{n}x_tw_t&=(1-\phi B)(1-B)Y_{T+1}-\phi(1-\phi B)(1-B)Y_{T+2}\\&=(1-\phi B)(1-\phi F)(1-B)Y_{T+1}\\&=[(1+\phi+\phi^2)Y_{T+1}-\phi Y_{T+2}]-[(1+\phi+\phi^2)Y_T-\phi Y_{T-1}]\end{aligned}$$

于是对这种情形我们有

$$\hat{\omega}=(1+\phi^2)^{-1}\{[(1+\phi+\phi^2)Y_{T+1}-\phi Y_{T+2}]-[(1+\phi+\phi^2)Y_T-\phi Y_{T-1}]\} \tag{12.1.13}$$

方差为 $\mathrm{var}[\hat{\omega}]=\sigma_a^2/(1+\phi^2)$ 。同样，估计值 $\hat{\omega}$ 仍可视作两个相同形式的加权平均之间的差别，其中一个是干预后的观测值 Y_{T+1} 和 Y_{T+2} , 另一个是干预前的观测值 Y_T 和 Y_{T-1}, 不过加权平均的阶数是有限的。这是因为噪声模型中仅有 AR 因子 $(1-\phi B)$, 而不象前面的情形那样有 MA 因子。

最后，我们考虑模型 (12.1.9) 的一个更简单情形 — 其中的噪声是平稳的，比如考虑 AR(1) 模型 $(1-\phi B)N_t = a_t$ 。这时我们得到 $x_t = (1-\phi B)BS_t^{(T)} = 1-\phi$, 对于 $t > T+1, x_{T+1} = 1$ 及 $w_t = (1-\phi B)Y_t = Y_t - \phi Y_{t-1}$ 。于是很容易地得到

$$\hat{\omega} \simeq \frac{(1-\phi)\sum\limits_{t=T+1}^{n}(Y_t - \phi Y_{t-1})}{(n-T)(1-\phi)^2} \simeq \overline{Y}_2 \tag{12.1.14}$$

其中 $\overline{Y}_2 = (n-T)^{-1}\sum\limits_{t=T+1}^{n} Y_t$ 记干预之后所有观测值的等权平均，方差为 $\text{var}[\hat{\omega}] = \sigma_a^2/[1+(n-T-1)(1-\phi)^2] \simeq \sigma_a^2/[(n-T)(1-\phi)^2]$。注意，由于噪声的平稳性，我们得到的是干预之后的简单平均，而且对干预之前的观测值也未加调整，这是因为假设了它们具有已知的零均值，且是平稳的。还需要注意的是：在平稳的情形，方差 $\hat{\omega}$ 依比例 $1/(n-T)$ 下降，而在前面非平稳的情形，方差 $\text{var}[\hat{\omega}]$ 基本上是常数，不依赖于样本的大小。这反映了干预模型参数估计值精度的差异程度，例如，在大样本情形，在平稳噪声模型和非平稳噪声模型之间，水平位移参数 ω 的差异是可以想见的。

具体地，在模型 (12.1.9) 中，$\xi_t = BS_t^{(T)}$ 是阶跃输入，假设噪声过程 N_t 是 $d=1$ 的非平稳 ARIMA，有 $\phi(B)(1-B)N_t = \theta(B)a_t$ 。于是，利用差分算子 $(1-B)$，模型

$$Y_t = \omega BS_t^{(T)} + N_t \tag{12.1.15}$$

还可以表示为

$$y_t = \omega BP_t^{(T)} + n_t \tag{12.1.16}$$

其中 $y_t = (1-B)Y_t$ 及 $n_t = (1-B)N_t$，从而 n_t 是平稳 ARMA(p,q) 过程。因此原模型 (12.1.15) 和模型 (12.1.16) 中 ω 的 MLE(极大似然估计) 具有相似的特征，前者是 (持续的) 阶跃影响和非平稳噪声 $(d=1)$，而后者是 (瞬时的) 脉冲输入和平稳噪声。

当然，模型 (12.1.9) 可以推广到允许在干预之前有一个未知的非零均值 ω_0，即 $Y_t = \omega_0 + \omega\xi_t + N_t$，其中 $\xi_t = BS_t^{(T)}$，因此，

ω 表示在干预之后均值水平的变化。那么，对于平稳 AR(1) 噪声模型的情形 (例如，类似于 (12.1.14))，可以证明 ω 的 MLE 是 $\hat{\omega} \simeq \overline{Y}_2 - \overline{Y}_1$，其中 $\overline{Y}_1 = T^{-1}\sum_{t=1}^{T} Y_t$ 表示干预之前所有观测值的样本均值。

12.2 时间序列的异常值分析

时间序列的观测值有时会受异常事件、干扰或误差的影响，它们会造成虚假的后果，导致观测值的反常态势，以致与时间序列中大多数观测值不一致。这些非正常的观测值可以称之为异常值，它们可以由非正常的外部事件引起，如：罢工，突发性的政治或经济变化，物理系统的突变，等等，或者仅仅是由于观测中的记录或采集的错误。在时间序列中异常值的存在会对样本自相关、偏相关、ARMA 模型参数的估计、预测等产生极大的影响，甚至可能影响到模型的识别。如果引起异常行为的干预事件发生的时刻 T 是已知的，那么，其异常影响可以用节 12.1 中讨论的干预分析技术来考虑。但是，由于在实际中着手进行分析时，异常值的出现常常是未知的，因此，检测异常值并估计其可能影响的辅助处理就变得十分重要。在本节中，我们讨论描述异常值的一些有用的模型及相应的方法，与干预分析类似的一些方法可用于异常值的检测。有关异常值检测、异常值的影响及估计的稳健方法等研究课题可见新近的参考文献 [67],[69],[144] 和 [190],

12.2.1 有关附加异常值和新息异常值的模型

根据 Fox[92] 的工作，我们考虑两种简单的干预模型，用以描述在实际中可能出现的两类异常值。它们是附加异常值 (AO) 和新息异常值 (IO) 模型。令 z_t 记不受异常值影响的基本时间序列过程，Y_t 记观测到的时间序列。我们假设 z_t 服从 ARIMA(p,d,q) 模型：　$\varphi(B)z_t = \theta(B)z_t$ 。那么，在时刻 T 的附加异常值 (或观

测到的异常值) 可以用模型表示为

$$Y_t = \omega P_t^{(T)} + z_t = \omega P_t^{(T)} + \frac{\theta(B)}{\varphi(B)} a_t \tag{12.2.1}$$

这里 $P_t^{(T)} = 1$, 若 $t = T; P_t^{(T)} = 0$, 若 $t \neq T$, 用来表示在时刻 T 的示性脉冲。在时刻 T 的新息异常值 (IO)(或新息冲击) 可以用模型表示为

$$Y_t = \frac{\theta(B)}{\varphi(B)}(\omega P_t^{(T)} + a_t) = \omega \frac{\theta(B)}{\varphi(B)} P_t^{(T)} + z_t \tag{12.2.2}$$

因此，一个附加异常值 (AO) 只影响被观测序列的水平，且是通过未知的附加值 ω 而起作用的： $Y_T = \omega + z_T$, 而新息异常值表示在时刻 T 的非正常的随机冲击： $a_T + \omega = a_T^*$, 它通过由 $\theta(B)/\varphi(B)$ 所描述的系统动态特性而影响后面的所有观测 $Y_T, Y_{T+1}, \cdots$ 。更一般地，一个被观测时间序列 Y_t 可以在多个时间点 $T_1, T_2, \cdots, T_k$ 上受到不同类型异常值的影响，那么，可以考虑多个异常值的模型，其一般形式为

$$Y_t = \sum_{j=1}^{k} \omega_j v_j(B) P_t^{(T_j)} + z_t \tag{12.2.3}$$

其中，对于时刻 T_j 的 AO 有 $v_j(B) = 1$; 对于时刻 T_j 的 IO 有 $v_j = \theta(B)/\varphi(B)$ 。与这些异常值模型有关的有意义的问题是：识别异常值发生的时间和种类，估计异常值影响的大小 ω, 据此分析时间序列并对异常值的影响作出调整。

12.2.2 发生时间已知的异常值影响估计

首先，对于 (12.2.1) 中的 AO 和 (12.2.2) 中的 IO, 我们分别考虑对影响 ω 的估计，这时，我们假设基本过程 z_t 的时间序列模型参数都是已知的。为了使用在一般情形下给出的迭代过程，还要假设异常值发生的时刻 T 是给定的。

令 $\pi(B)=\theta^{-1}(B)\varphi(B)=1-\sum_{i=1}^{\infty}\pi_i B^i$，并定义 $e_t=\pi(B)Y_y$，$t=1,2,\cdots,n$，观测序列为 Y_t。于是我们可以把异常值模型分别写成

$$\text{IO}:\quad e_t=\omega P_t^{(T)}+a_t \tag{12.2.4a}$$

$$\text{AO}:\quad e_t=\omega\pi(B)P_t^{(T)}+a_t=\omega x_{1t}+a_t \tag{12.2.4b}$$

这里，对于 AO 模型，$x_{1t}=\pi(B)P_t^{(T)}=-\pi_i$，若 $t=T+i>T$；$x_{1t}=0$，若 $t<T$，有 $\pi_0=-1$。这样，从 (12.2.4) 中可看到关于 IO 的信息仅包含在具体时刻 T 的"残差"e_T 之中，而 AO 的信息则分散到残差 $e_T,e_{T+1},e_{T+2},\cdots$ 中，并具有递减的权数 $1,-\pi_1,-\pi_2,\cdots$，这是因为由 MA 算子 $\theta(B)$ 的可逆性，π_i 是绝对可和的。

根据最小二乘原理，IO 模型中异常值影响 ω 的最小二乘估计就是时刻 T 的残差

$$\text{IO}:\quad \hat{\omega}_{I,T}=e_t \tag{12.2.5a}$$

其中 $\operatorname{var}[\hat{\omega}_{I,T}]=\sigma_a^2$，我们注意到在 AO 模型中，异常值的影响则是 $e_T,e_{T+1},\cdots$ 的线性组合：

$$\text{AO}:\quad \hat{\omega}_{A,T}=\frac{e_T-\sum_{i=1}^{n-T}\pi_i e_{T+i}}{\sum_{i=0}^{n-T}\pi_i^2}=\frac{\pi^*(F)e_T}{\tau^2} \tag{12.2.5b}$$

其中 $\operatorname{var}[\hat{\omega}_{A,T}]=\sigma_a^2/\tau^2$，这里 $\tau^2=\sum_{i=0}^{n-T}\pi_i^2$，$\pi^*(F)=1-\pi_1F-\pi_2F^2-\cdots-\pi_{n-T}F^{n-T}$。在 (12.2.5) 的表示中反映了这些估计量依赖于时刻 T 这一事实。注意，在基本自回归模型 $\varphi(B)z_t=a_t$ 中，由于对 $T<n-p-d$ 有 $\pi^*(B)=\pi(B)=\varphi(B)$，及 $e_t=\varphi(B)Y_t$，则 (12.2.5b) 中的估计值 $\hat{\omega}_{A,T}$ 用观测序列 Y_t 表示，可以写成

$$\hat{\omega}_{A,T}=\frac{\varphi(F)\varphi(B)Y_T}{\tau^2}$$

由于 $\tau^2 > 1$, 可知一般有 $\mathrm{var}[\hat{\omega}_{A,T}] < \mathrm{var}[\hat{\omega}_{I,T}] = \sigma_a^2$, 有时 $\mathrm{var}[\hat{\omega}_{A,T}]$ 可以比 σ_a^2 小很多。例如，在 z_t 为 MA(1) 的模型中，当 $n-T$ 很大时，$\hat{\omega}_{A,T}$ 的方差为 $\sigma_a^2(1-\theta^2)/(1-\theta^{2(n-T+1)}) \simeq \sigma_a^2(1-\theta^2)$, 就是这种情形。

在给定时刻 T 给出的 AO 型或 IO 型异常值存在与否的显著性检验，可以转化为检验模型 (12.2.1) 或 (12.2.2) 中是 $\omega = 0$ 还是 $\omega \neq 0$ 。两种情形下都可以导出似然比检验准则，对于 IO 型和 AO 型，标准统计量的形式分别为

$$\lambda_{I,T} = \frac{\hat{\omega}_{I,T}}{\sigma_a} \quad \text{和} \quad \lambda_{A,T} = \frac{\tau\hat{\omega}_{A,T}}{\sigma_a} \tag{12.2.6}$$

在零假设 $\omega = 0$ 之下，在 (12.2.6) 中的两个统计量都具有标准正态分布。

12.2.3 异常值检验的迭代方法

在实际中，异常值可能出现的时间和模型参数都是未知的。为了解决在未知时刻异常值的检测问题，Chang,Tiao 和 Chen[69] 提出了计算起来较简便的迭代方法，以识别和修正异常值的影响。该方法的第一步是：假设序列不含异常值，用常规方法对观测到的序列 Y_t 估计 ARIMA 模型。模型的残差 $\hat{e}_t$ 由 $\hat{e}_t = \hat{\theta}^{-1}(B)\hat{\varphi}(B)Y_t = \hat{\pi}(B)Y_t$ 得到，且 $\hat{\sigma}_a^2 = n^{-1}\sum_{t=1}^{n}\hat{e}_t^2$ 。于是可对每个时刻 $t = 1, 2, \cdots, n$ 计算 (12.2.6) 中的统计量 $\hat{\lambda}_{I,t} = \hat{\omega}_{I,t}/\hat{\sigma}_a$ 和 $\hat{\lambda}_{A,t} = \hat{\tau}\hat{\omega}_{A,t}/\hat{\sigma}_a$, 以及 $\hat{\lambda}_T = \max_t[\max(|\hat{\lambda}_{I,t}|, |\hat{\lambda}_{A,t}|)]$, 这里用 T 记最大值发生的时刻。如果 $\hat{\lambda}_T = |\hat{\lambda}_{I,T}| > 0$, 则识别出了在时刻 T 可能有 IO 型异常值，这里 c 是预先确定的常数，c 的典型值为 3.0,3.5 或 4.0. 在 T 时刻，通过定义 $\tilde{e}_T = \hat{e}_T - \hat{\omega}_{I,T} = 0$ 可以把这一 IO 的影响从残差中消除。如果 $\hat{\lambda}_T = |\hat{\lambda}_{A,T}| > c$, 则识别出在 T 时刻可能有 AO 型的异常值，其影响由 (12.2.5b) 中的 $\hat{\omega}_{A,T}$ 来估计。通过对 $t > T$ 定义 $\tilde{e}_t = \hat{e}_t - \hat{\omega}_{A,T}\hat{\pi}(B)P_{t-T}^{(T)} = \hat{e}_t + \hat{\omega}_{A,T}\hat{\pi}_{t-T}$, 可以将这一 AO 型异

常值的影响从残差中移去。无论是哪种情形，新的估计值 $\hat{\sigma}_a^2$ 都可以由修正后的残差 $\widetilde{e}_t$ 计算出。

当识别出任何异常值后，应使用修正的残差 $\widetilde{e}_t$ 和修正的估计量 $\hat{\sigma}_a^2$，但参数 $\hat{\pi}(B)=\hat{\theta}^{-1}(B)\hat{\varphi}(B)$ 还是用原先的，这样就可以计算出新的统计量 $\hat{\lambda}_{I,t}$ 和 $\hat{\lambda}_{A,t}$。然后重复前面的步骤，直到识别出所有的异常值。假设由此方法识别出在 k 个时点 $T_1,T_2,\cdots,T_k$ 有异常值，然后对观测序列 Y_t 估计形如 (12.2.3) 的异常值的总模型

$$Y_t=\sum_{j=1}^{k}\omega_j v_j(B)P_t^{(T_j)}+\frac{\theta(B)}{\varphi(B)}a_t \qquad (12.2.7)$$

其中，在时刻 T_j，对于 AO 型有 $v_j(B)=1$，而对于 IO 型有 $v_j=\theta(B)/\varphi(B)$。由这个拟合模型可以得到一组修正的残差

$$\hat{e}_t=\hat{\theta}^{-1}(B)\hat{\varphi}(B)[Y_t-\sum_{j=1}^{k}\hat{\omega}_j\hat{v}_j(B)P_t^{(T_j)}]$$

及新的 $\hat{\sigma}_a^2$。然后，这一方法的前面步骤可以用新的残差重复下去，直到识别出所有异常值，并估计出 (12.2.7) 的一般形式的最终模型。如果需要的话，从观测值中剔除异常值影响的修正时间序列可由 $\widetilde{z}_t=Y_t-\sum_{j=1}^{k}\hat{\omega}_j\hat{v}_j(B)P_t^{(T_j)}$ 构造出。以上方法几乎不用修改就可以补充到现有的估计 ARIMA 和传递函数 — 噪声模型的软件中。这种技巧对于识别时间序列隐含的异常值是很有用的工具，如果这些异常值未经检测，会对建模和估计的有效性产生负面的影响。但是，从头到尾完整地使用这种“异常值调整”方法时应该多加小心，尤其要注意结果的整体解释，以及一般模型对“异常值”的假设是否合理。例如，在 (12.2.7) 中将异常值处理为确定常数是否恰当，过多设定异常值数目的可能性是否存在。如果有可能，最好对那些用前面方法识别出来的异常值找出其原因或根源，这样才能更好地弄清异常行为，并在分析中正确地解释。

12.2.4 异常值分析的例子

为了说明前几节讨论的异常值分析方法的应用，我们考虑两个数值例子。为了计算简便，在这些例子中都使用条件最小二乘估计方法。

序列 D　第一个例子涉及到序列 D，它表示从一个化学过程每小时读到的“非控制”粘度。在第 7 章中曾提出用一个 AR(1) 模型 $(1-\phi B)z_t=\theta_0+a_t$ 来描述，并作了拟合。在异常检测方法中，首先在不存在异常值的假设之下估计模型，其结果在表 12 .1(a) 中给出。然后，利用 $\hat{\sigma}_a^2=0.08949$，对每个时刻 t 计算 (12.2.6) 中的 AO 和 IO 统计量。由临界值 $c=3.5$，我们识别出在时刻 $T=217$ 有一个很大的 IO 。其影响可以通过修正 T 时刻的残差而去掉，于是得到新的估计 $\tilde{\sigma}_a^2=0.08414$ 。利用 $\tilde{\sigma}_a$ 可以计算出新的异常值统计量。在这一阶段中没有异常值被识别出。然后估计时间序列参数和模型 (12.2.2) 中的异常值参数 ω，即在下面的模型

$$Y_t=\frac{1}{1-\phi B}[\theta_0+\omega P_t^{(T)}+a_t]$$

表 12.1 对序列 C 和 D 的异常值检验和参数估计结果

	参数 *						异常值			
	$\hat{\theta}_0$	$\hat{\phi}$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\sigma}_2^2$	时间	$\hat{\omega}$	$\hat{\lambda}$	类型
(a) 序列D										
第一轮	1.269	0.862				0.0895	217	−1.28	−4.29	IO
	(0.258)	(0.028)								
终值	1.181	0.872	−1.296			0.0841				
	(0.251)	(0.027)	(0.292)							
(b) 序列C										
第一轮		0.813				0.0179	58	0.76	5.65	IO
		(0.038)					59	−0.51	−4.16	IO
							60	−0.44	−3.47	IO
终值		0.851	0.745	−0.551	−0.455	0.0132				
		(0.035)	(0.116)	(0.120)	(0.116)					

* 在括号中是参数估计值的标准差。

中，同时将它们估计出。估计值在表 12.1(a) 中给出。根据这些新的参数估计值和相应的残差，重复异常值检测过程，没有发现其它的异常值，因此最终只有一个 IO 型异常值被识别出来。调整这个异常值对于该时间序列模型参数估计值 $\hat{\phi}$ 的变化影响不大，但使 σ_a^2 的估计减少了大约 6% 。在异常值的检验过程中，时刻 t =29,113,115,171,268 和 272 的检验统计量 $\hat{\lambda}$ 的值按绝对值稍大于 3.0, 这意味着这几个值可能是异常值，但是对这几个值的调整对于模型的估计没有显著的影响。

序列 C　我们考虑的第二个例子是序列 C, 它是从一个化工生产中每分钟读到的“未控制”温度。前文中对这个序列识别和拟合的模型是 ARIMA(1,1,0): $(1-\phi B)(1-B)\dot{z}_t = a_t$ 。假设无异常值时对于该模型得到的估计结果在表 12.1(b) 中给出。通过对异常值检验统计量的一系列计算，并取临界值 $c=3.5$, 我们首先识别出在时刻 58 的 IO 异常值. 对时刻 58 的残差修正后，我们得到新的估计值 $\tilde{\sigma}_a^2$ =0.01521, 下一个识别出的 IO 异常值得是在时刻 59. 修正残差，得到新的估计值 $\tilde{\sigma}_a^2$ =0.01409, 于是在时刻 60 又识别出另一个 IO 异常值。此后，再没有识别出其它异常值。图 12.2 给出了在异常值调整之前，用初始模型拟合所得前 100 个残差的时间序列图，从图中看到，在时刻 58,59 和 60 的新息异常值是非常明显的。

对该序列估计出的时间序列异常值模型是

$$(1-B)Y_t = \frac{1}{1-\phi B}[\omega_1 P_t^{(58)} + \omega_2 P_t^{(59)} + \omega_3 P_t^{(60)} + a_t]$$

结果在表 12.1(b) 中给出。对新估计的模型参数重复异常值识别过程，再没有异常值被检测出。在这个例子中我们看到，异常值调整对于时间序列参数估计值 $\hat{\phi}$ 的影响略大于前面的例子，并使 σ_a^2 的估计值有显著地下降 (约为 26%) 。

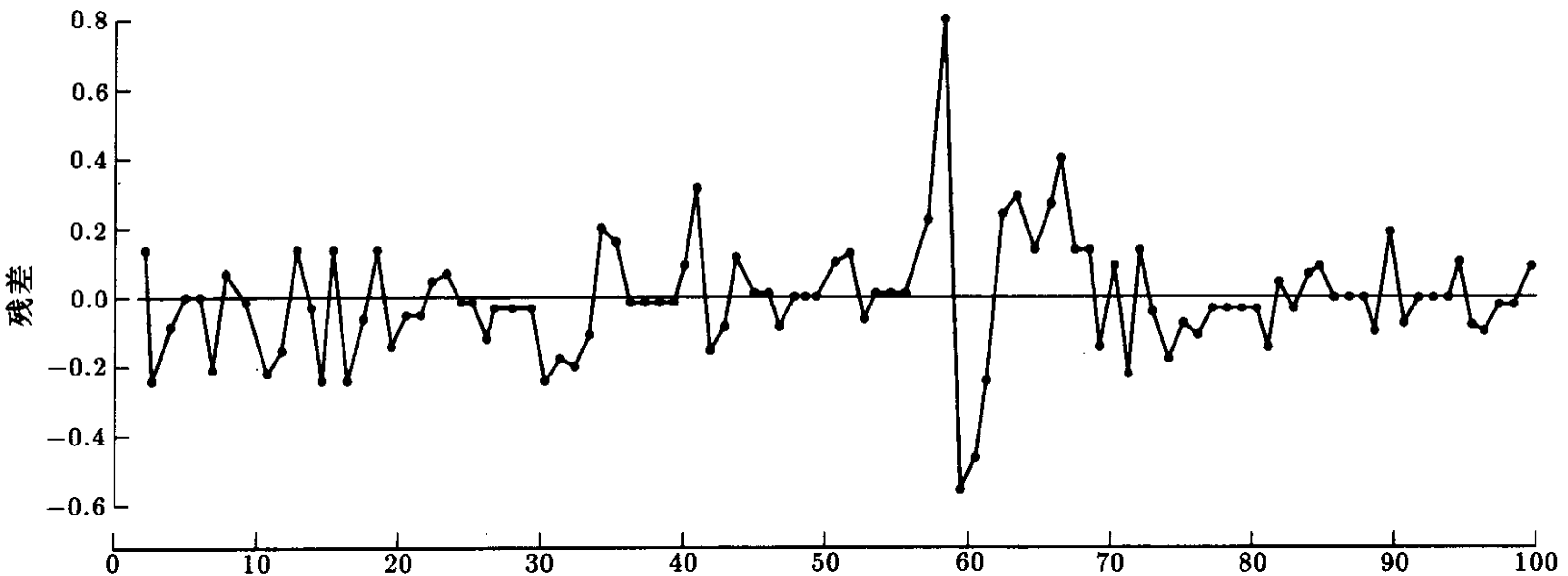


图 12.2　对序列 C 拟合 ARIMA(1,1,0) 模型的前 100 个残差值

12.3 对存在缺失值 ARMA 模型的估计

在实际中的许多情形下，由于对应于某些时间点有“缺失值”，因此时间序列 z_t 并不是在等间隔的时间上被观测到。在本节中，我们通过考虑计算观测数据的精确似然函数，简要地讨论这种情形下 ARIMA(p,d,q) 模型的极大似然估计。可以证明，在缺失数据的情况下，如同在 5.5 和 5.7 节中已讨论的，利用模型的状态空间形式及相关的 Kalman 滤波方法可以容易且方便地构造出似然函数，但要加以修改使之适应于缺失数据。有关对不规则间隔的观测值计算似然函数的方法，对此问题曾做过研究的有 Jones[123],Harvey 和 Pierse[108],Ansley 和 Kohn[12],[13], 及 Wincke 和 Riensel[205]。

我们假设在不等间隔的整数时间 $t_1 < t_2 < \cdots < t_n$ 上得到 n 个观测值，它们来自 ARIMA(p,d,q) 过程，服从于模型 $\phi(B)(1-B)^d z_t = \theta(B)a_t$. 由 5.5 节，过程 z_t 具有的状态空间模型形式由下式给出

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}\boldsymbol{Y}_{t-1} + \boldsymbol{\Psi}a_t \tag{12.3.1}$$

其中 $z_t = \boldsymbol{H}\boldsymbol{Y}_t = [1,0,\cdots,0]\boldsymbol{Y}_t$, 这里 $\boldsymbol{Y}_t$ 是 r 维状态向量， $r = \max(p+d, q+1)$ 。令 $\Delta_i = t_i - t_{i-1}$ 记顺次观测值 z_{t_i-1} 和 $z_{t_i}, i = 2,\cdots,n$ 之间的时间差。在 (12.3.1) 的右边经过 Δ_i 次替代，我们得到

$$\boldsymbol{Y}_{t_i} = \boldsymbol{\Phi}^{\Delta_i}\boldsymbol{Y}_{t_{i-1}} + \sum_{j=0}^{\Delta_i-1}\boldsymbol{\Phi}^j\boldsymbol{\Psi}_{t_i-j} \equiv \boldsymbol{\Phi}_i^*\boldsymbol{Y}_{t_{i-1}} + \mathbf{a}_{t_i}^* \tag{12.3.2}$$

其中 $\boldsymbol{\Phi}_i^* = \boldsymbol{\Phi}^{\Delta_i}$ 及 $\mathbf{a}_t^* = \sum_{j=0}^{\Delta_i-1}\boldsymbol{\Phi}^j\boldsymbol{\Psi}a_{t_i-j}$, 且有

$$\mathrm{cov}[\mathbf{a}_{t_i}^*] = \Sigma_i = \sigma_a^2\sum_{j=0}^{\Delta_i-1}\boldsymbol{\Phi}^j\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{\Phi}'^j$$

于是， (12.3.2) 和量测方程 $z_{t_i} = \boldsymbol{H}\boldsymbol{Y}_{t_i}$ 一起构成了被观测时间序列数据 $z_{t_1}, z_{t_2},\cdots,z_{t_n}$ 的状态空间模型表示。

因此，如 (5.5.5)—(5.5.8) 的 Kalman 滤波递推公式可以值接用来得到状态预报 $\hat{\boldsymbol{Y}}_{t_i|t_i-1}$ 和相应的误差协方差阵 $\boldsymbol{V}_{t_i|t_i-1}$。于是我们由 Kalman 滤波递推算法，很容易得到基于过去观测数据 z_{t_i} 的预报值

$$\hat{z}_{t_i|t_{i-1}} = E[z_{t_i}|z_{t_{i-1}}, \cdots, z_{t_i}] = \boldsymbol{H}\hat{\boldsymbol{Y}}_{t_i|t_{t-1}} \tag{12.3.3}$$

及误差协方差阵

$$\sigma_a^2 v_i = \boldsymbol{H}\boldsymbol{V}_{t_i|t_{i-1}}\boldsymbol{H}' = E[(z_{t_i} - \hat{z}_{t_i|t_{i-1}})^2] \tag{12.3.4}$$

更具体地说，在缺失数据情形 (5.5.5) 和 (5.5.6) 的修正公式为如下形式

$$\hat{\boldsymbol{Y}}_{t_i|t_i} = \hat{\boldsymbol{Y}}_{t_i|t_{i-1}} + \boldsymbol{K}_i(z_{t_i} - \boldsymbol{H}\hat{\boldsymbol{Y}}_{t_i|t_{i-1}}) \tag{12.3.5}$$

其中

$$\boldsymbol{K}_i = \boldsymbol{V}_{t_i|t_{i-1}}\boldsymbol{H}'[\boldsymbol{H}\boldsymbol{V}_{t_i|t_{i-1}}\boldsymbol{H}']^{-1} \tag{12.3.6}$$

而预报方程 (5.5.7) 由下式给出

$$\begin{aligned} \hat{\boldsymbol{Y}}_{t_i|t_{i-1}} &= \boldsymbol{\Phi}_i^*\hat{\boldsymbol{Y}}_{t_{i-1}|t_{i-1}} = \boldsymbol{\Phi}^{\Delta_i}\hat{\boldsymbol{Y}}_{t_{i-1}|t_{i-1}} \\ \boldsymbol{V}_{t_i|t_{i-1}} &= \boldsymbol{\Phi}_i^*\boldsymbol{V}_{t_{i-1}|t_{i-1}}\boldsymbol{\Phi}_i^{*\prime} + \Sigma_i \end{aligned} \tag{12.3.7}$$

其中

$$\boldsymbol{V}_{t_i|t_i} = [\boldsymbol{I} - \boldsymbol{K}_i\boldsymbol{H}]\boldsymbol{V}_{t_i|t_{i-1}} \tag{12.3.8}$$

注意，计算预报方程 (12.3.7) 可以解释为对 $j = 1, \cdots, \Delta_i$ 顺次计算提前 1 期预报：

$$\begin{aligned} \hat{\boldsymbol{Y}}_{t_{i-1}+j|t_{i-1}} &= \boldsymbol{\Phi}\hat{\boldsymbol{Y}}_{t_{i-1}+j-1|t_{i-1}} \\ \boldsymbol{V}_{t_{i-1}+j|t_{i-1}} &= \boldsymbol{\Phi}\boldsymbol{V}_{t_{i-1}+j-1|t_{i-1}}\boldsymbol{\Phi}' + \sigma_a^2\boldsymbol{\Psi}\boldsymbol{\Psi}' \end{aligned}$$

在这里不需要作任何修正，因为在时刻 t_{i-1} 和 t_i 之间没有观测值来为修正提供另外的信息。

观测向量 $\boldsymbol{z}'=(z_{t_1},z_{t_2},\cdots,z_{t_n})$ 的精确似然函数可由 (12.3.3) 和 (12.3.4) 中的量得到，这是因为 $\boldsymbol{z}$ 的联合密度可表示为当 $z_{t_{i-1}}$, $\cdots,z_{t_1}$ 给定时 z_{t_i} 的条件密度的乘积，这里 $i=2,\cdots,n$)，这些条件密度都是 Gauss 型的，(条件) 均值和方差由 (12.3.3)-(12.3.4) 给出。所以观测 $\boldsymbol{z}$ 的联合密度可表示为

$$p(\boldsymbol{z}|\boldsymbol{\phi},\boldsymbol{\theta},\sigma_a)=\prod_{i=1}^{n}(2\pi\sigma_a^2v_i)^{-1/2}\exp\left[-\frac{1}{2\sigma_a^2}\sum_{i=1}^{n}\frac{(z_{t_i}-\hat{z}_{t_i|t_{i-1}})^2}{v_i}\right] \tag{12.3.9}$$

在 (12.3.9) 中，量值 $\hat{z}_{t_i|t_{i-1}}$ 和 $\sigma_q^2v_i$ 可直接由递推滤波算式 (12.3.5) -(12.3.8) 确定。对于平稳 ARMA(p,q) 模型的情形，着手滤波递推算法所需要的初始条件很容易决定 (如 Jones[123])，但是对于非平稳 ARIMA 模型的情形，则需要对过程和初始条件设定其它假设。有关这种情形的适用方法 Ansley 和 Kohn[13] 曾作过研究。

为了说明处理缺失数据的方法，作为一个简单的例子，我们考虑平稳 AR(1) 模型 $(1-\Phi B)z_t=a_t$，这时 (12.3.2) 直接写成 (例如，可见 [168])

$$z_{t_i}=\phi^{\Delta_i}z_{t_{i-1}}+\sum_{j=0}^{\Delta_i-1}\phi^j a_{t_i-j} \tag{12.3.10}$$

容易确定出

$$\hat{z}_{t_i|t_{i-1}}=\phi^{\Delta_i}z_{t_{i-1}} \quad 及 \quad \sigma_i^2=\sigma_a^2v_i=\frac{\sigma_a^2(1-\phi^{2\Delta_i})}{1-\phi^2} \tag{12.3.11}$$

因此，利用这些 $\hat{z}_{t_i|t_{i-1}}$ 和 $\sigma_a^2v_i$ 的表达式，(12.3.9) 给出了有缺失值的一阶自回归模型观测数据的似然函数。

ARMA 过程缺失值的估计　在时间序列缺失值的有关内容中，常会引起一个有意义的相关问题，即对缺失值的估计。早先

对 ARIMA 时间序列缺失值的插值问题是从最小二乘的角度加以考察，可见 Brubacher 和 Tunnicliffe Wilson[66],Damsleth[75]，以及 Abraham[1]。在状态空间表示的框架中，缺失值的估计及相应的误差方差可以方便地用与 Kalman 滤波相联系的递推平滑方法导出，比如 Anderson 和 Moore[5] 所描述的一般形式。Harver 和 Pierse[108], 以及 Kohn 和 Ansley[129] 更具体地对 ARIMA 模型缺失数据的情形讨论了一些方法。

但是，对于纯自回归模型 $\phi(B)z_t = a_t$, 可以得到一些较为简单的显式插值结果. 例如，在一个 AR(p) 过程中，若只在时刻 T 有一个缺失值，T 时刻之前和之后都至少有 p 个连续的观测值; 众所周知 (如见 [66]), 缺失值 z_T 的最优插值由下式给出

$$\hat{z}_T = -d_0^{-1}\sum_{j=1}^{p} d_j(z_{T-j} + z_{T+j}) \tag{12.3.12}$$

其中 $d_j = \sum_{i=j}^{p}\phi_i\phi_{i-j}, \phi_0 = -1$, 及 $d_0 = 1+\sum_{i=1}^{p}\phi_i^2$, 这里 $E[z_T - \hat{z}_T)^2] = \sigma_0^2 d_0^{-1} = \sigma_a^2(1+\sum_{i=1}^{p}\phi_i^2)^{-1}$。注意，(12.3.12) 中的值可以表示为 $\hat{z}_T = z_T - [\phi(B)\phi(F)z_T/d_0]$, 插值误差等于

$$\hat{e}_T = z_T - \hat{z}_T = \frac{\phi(B)\phi(F)z_T}{d_0} \tag{12.3.13}$$

在一般 ARMA 模型的情形，包括 Bruce 和 Martin[67] 在内的许多人都注意到：缺失值情形的似然函数结构与具有 AO 的连贯数据的似然函数表示有着密切的联系，只需在每一个缺失值对应的时点上都将其确定为附加异常值 (AO)。因此实际上，在这样一个具有连贯数据的时间序列附加异常值模型中，异常值影响参数 ω 的估计对应于缺失数据情形的插值误差。例如在自回归模型的情形，可将 (12.3.13) 的结果和由 (12.2.5b) 给出的 AO 模型的结果作比较。而且，缺失数据情形似然函数 (12.3.9) 中的平方和函数等于对全部连贯观测得出的平方和，且似然函数的求取是

在如下情形：在出现缺失值的每一个时点上都假定有一个附加异常值，而且，当时间序列模型参数值 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 给定时，每个对应 AO 的异常影响参数 ω 都是极大似然估计。为了加以解释，我们用 AR(1) 模型的简单情形加以说明，在时刻 T 有一个单个孤立的缺失值，由 (12.3.11) 在缺失数据平方和函数中的相关项为

$$\begin{aligned}\frac{(Z_{T+1}-\phi^2 z_{T-1})^2}{1+\phi^2} &\equiv [(z_T-\hat{\omega})-\phi z_{T-1}]^2+[z_{T+1}-\phi(z_t-\hat{\omega})]^2\\ &=(\hat{z}_T-\phi z_{T-1})^2+(z_{T+1}-\phi\hat{z}_T)^2\end{aligned} \tag{12.3.14}$$

其中

$$\hat{\omega}=z_T-\frac{\phi}{1+\phi^2}(z_{T-1}+z_{T+1})=z_T-\hat{z}_T$$

是 AO 模型 (12.2.1) 中异常值影响 ω 的极大似然估计，(12.3.14) 的后一式子表示在时刻 T 具有 AO 模型的情形下，连贯数据情形的平方和项。

第四部分

离散控制方案的设计

在前面几章中我们研究了离散时间序列和动态系统的建模。我们看到，一旦得到恰当的模型之后，如何由它们得出时间序列的预报以及描述动态系统的传递函数。但是，这些模型及其操作比这里所指出的应用具有更广泛的重要性。我们所提出的这些构想在一类广泛的随机动态系统的分析中具有重要意义，例如发生在经济，工程，商业和组织研究中的动态系统。

要对所有的应用都加以详细说明是不可能的，但是，本书中的理论和实例希望能帮助读者将一般方法适应于自己的具体问题。在这样做的过程中，我们已经讨论过的动态随机模型往往可以象建筑模块一样联结在一起，用来描述所研究的具体动态系统。类似于我们已经详细叙述过的，识别、估计和诊断检验技术也是建立模型所需要的。最后，递推计算和在预测这个一般性题目下考

虑的许多设想也会广泛地应用于对已拟合模型得出一些结果。我们将通过反馈控制方案的设计来说明进一步应用的可能性，并以此结束本书。通观第 13 章，这是前文讨论问题的集中演练，并在全新的应用中尽可能完全地给出详细结果，我们希望这是有价值的。

13

过程控制的各个方面

工业控制这个术语有不同的使用方式。在涉及离散“部件”制造的工业生产中通常使用的是 Shewhart 图和其它控制图，即所谓的统计过程控制 (SPC)。与此不同的是，各种形式的前馈和反馈控制则使用在尤其是生产过程和化学工业中，我们称为工程过程控制 (EPC)。由于工程过程控制常常是自动计算和实现的，所以这类控制有时又称为自动过程控制 (APC)。但是，实现调整所采用的方式只是为了使用方便，因此，我们不在这里使用这个术语。本章的目的是根据本书前面的讨论来深入地观察这些控制方法的统计内涵，并更好地理解它们的关系和目标。

我们首先考虑过程的监视，可使用 Shewhart 控制图，并将它与过程调整技术相对照。一般的调节问题就是通过操纵输入变量以保持输出变量接近某一目标值，从而得到反馈控制。我们首先用完全直观的方法来考虑这个问题，然后将其与前面讨论过的一些随机模型和传递函数模型联系起来，得出输出为最小均方误差 (MMSE) 的控制方案。这里引出的离散方案的讨论类似于工程中的比例 - 积分 (PI) 方案，我们要说明如何利用 PI 控制对人工调整过程设计出简单的图形。

事实表明，最小均方控制往往要求对输入量进行极其大的调整。因此，可以引入带有约束的“最优”方案，它所需要的调整小得多，而代价只是输出的均方误差有微小的上升。这些带约束

的方案一般不是 PI 方案，但是在某些重要场合可以证明，适当选择的 PI 方案可以很好地接近其特性。

尤其是在涉及部件制造的工业中，会有与调整生产过程相联系的固定代价，在某些情形还有与获得观测相联系的监视代价。因此，当我们对反馈控制讨论有限制的调整方案时，应使得在对应场合全部代价最小。最后，我们利用 Cuscore 图，对于参数值的可能变化考虑监视方案的一般方法。更一般的讨论可见附录和参考文献。

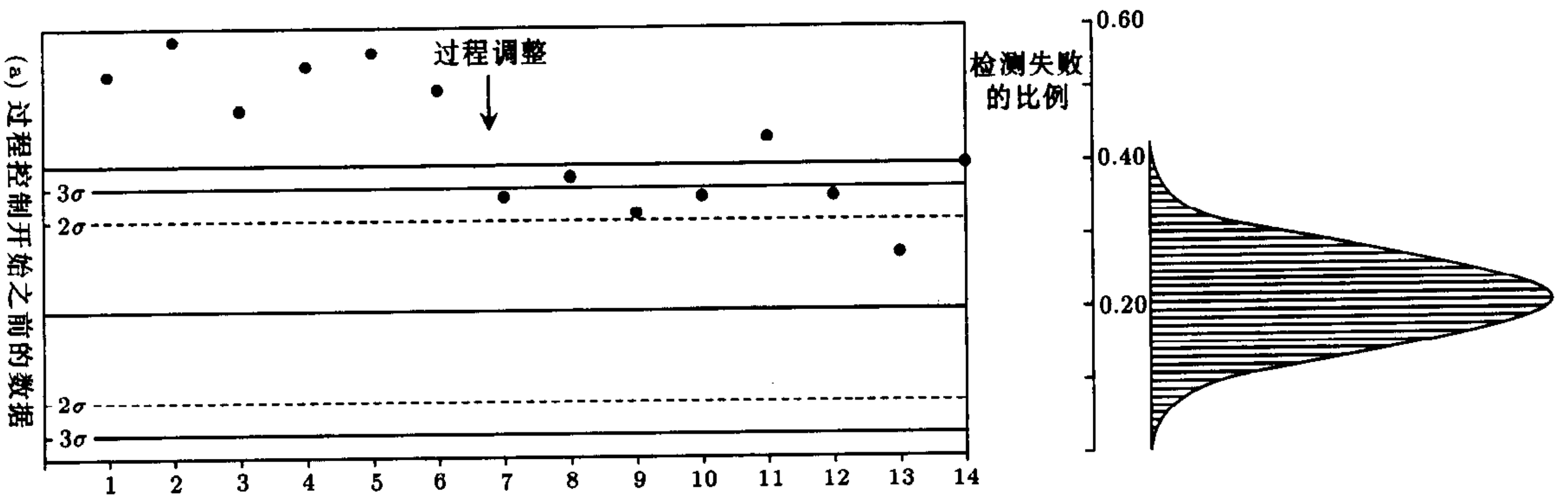
13.1 过程监视和过程调整

过程控制无非是力图消除基本物理定律 —热力学第二定律的影响，该定律指出：如果任其发展，任一系统的熵或无序性绝不会降低，而通常只会增加. 统计过程控制 (SPC) 和工程过程控制 (EPC) 是与这一定律抗争的两种互补方法。 SPA 试图用过程控制来消除干扰，而 EPC 试图用过程调整加以补偿 (又见 Box 和 Kramer[52]) .

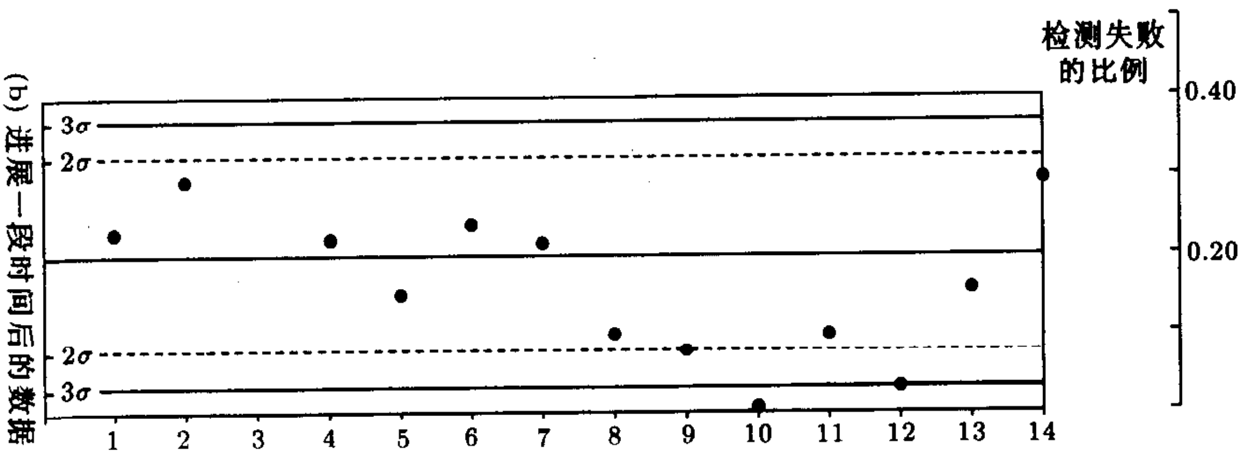
13.1.1 过程监视

SPC 为使过程稳定化的策略是：将生产流程和原材料标准化，利用能够设想并能实施的手段 (诸如图形，检验表， Parto 图，因果图等等) 来寻找并消除故障原因 (如可见 Ishikawa[116]) 由于寻找可指明的原因是枯燥而昂贵的，所以，通常要等到对稳定模型的“统计显著性 ” 偏离发生后再进行查找，这样做是合乎情理的。这些可通过使用过程监视图来实现，如 Shewart 图，Cusum 图和 Robert 的 EWMA 图。这里所奉行哲理是“没坏就别修 ”—不要无谓地调节过程 (如见 Deming[79]) .

图 13.1 给出使用 Shewhart 控制图进行过程监视的一个例子. 在避孕套的生产线上每隔 2 小时取 50 个产品来进行检验，



(a) 过程控制开始之前的数据



(b) 进展一段时间后的数据

图 13.1 避孕套充气检测失败比例的 Shewhart 图

将它们充气达到某个固定的高压，记下爆炸的比例。图 13.1 给出了启动生产这些产品的一台机器所取到的数据。由同类机器得出的研究表明，如果这一严格检验失败的比例为 $p = 0.20$, 则所生产的产品是高质量的。

图 13.1(a) 右边的棒图所示参考分布用来描绘理想的过程特征。这是一个二项分布，它给出了 p 为常数值 0.20 时在 n =50 的随机样本中得到失败比例的概率。如果数据的特征似乎是来自这一参考分布的随机序列，我们就说过程仍在控制状态下，不需要采取任何行动。与此相反，如果数据的状态不象这样，而是出现了异常值或可疑性状，我们就有理由认为发生了某些情况。在实际中，完整的参考分布通常是未知的，但可以设置上下控制线和警戒线。在这种情况下，当一条正态曲线 (以连续曲线出现) 为参考分布提供了很好的近似，则这些控制线通常设为 $\pm 2\sigma$ 和 $\pm 3\sigma$, 其中 $\sigma = \sqrt{p(1-p)/n}$ 是二项分布样本比例的标准差。在这个例子中，$p = 0.20, n = 50$, 可以得出 $\sigma = 0.057$ 。图 13.1(a) 说明在开始阶段，过程严重失控，未通过检测的比例高达 50% 。经 20 小时运行后所做的过程调整把次品数量降低到约 40%, 但仍需要进一步的改变，以使过程达到理想水平 $p = 0.20$ 的控制状态。通过一系列的处理行动，最终达到了这一目的，图 13.1(b) 给出了后一阶段进展中的过程运行。尽管该系统现在看来大概是处于理想控制状态，但是应注意到，图中第 10 个点落在 $\pm 3\sigma$ 线的下限之外. 接下来的研究表明，这一异常点是由检测过程引起的。在供气线上发生了一次故障，从而当时检测避孕套的气压大大降低，导致次品比例错误地降低。经过采取纠正行动对系统进行修正，使得检测机器只在气压正确设置时才能运行，这样就保证了这一具体错误不会再次发生。

这种监视过程显然是很有价值的。依 Shewhart[177] 和 Deming[79] 所述，我们称过程处于控制状态的偏离 (此例中 $n = 50, p = 0.20$ 时的二项偏离) 是由一般原因所致。存在一般原因的系统只能被那些影响系统的处理行动所改变。因此，如果可接受的次品

率成为 10% 时，可以引入一种新的检测机器。这时一般原因的偏离就是 $p = 0.10$ 的二项偏离。

通过使用图形而发现的供气线上的故障称之为特定[①]原因。经过适当的"检测"工作，设备操作者通常可以找到并消除特定原因。因此，过程监视的目的是：(1) 建立并持续地保证通常的理想状态使系统维持运行，(2) 寻找那些不大可能是偶发的偏离，它们可以引导跟踪并消除出现故障的可指明原因。

13.1.2 过程调整

尽管我们必须经常地作出不懈的努力来消除偏差的原因，比如，令人不满意的检测方法，原材料的差异，操作员的差异，等等。但通过这种方法仍不能使某些过程达到满意的稳定状态。即使我们尽了最大的努力，过程总还是有偏离目标的趋势。这可归因于那些已知但不可控的现象，比如周围温度、湿度的变化，原材料的质量，等等，或者是当前尚未知的原因。在这种情况下，一些对过程的调整或对系统的校正是必要的。在这里通过对某些附加变量的操作，来补偿质量特征的偏差。

为了确定思想，我们首先在纯经验的基础上介绍一个简单的反馈方案，理论上的合理性留待以后。考虑图 13.2 所示的测量数据，这是在等间隔时间上所取的某种很薄的金属膜的厚度。其质量特征已严重失控，但标准的处理方法又不能使它稳定下来(Box[36])。假设扰动 N_t 定义为没有进行调整时质量特征对其目标值的偏离，也即 N_t 是隐含的噪声过程。在假设有一个可控变量 —沉淀速度 X— 可以方便地用于调节厚度，每一单位的 X 变化会引起厚度 g 个单位的变化，并且在一个时间间隔内便能完全生效。如果在 t 时刻令 X 等于 X_t，那么，在 $t+1$ 时刻自目标值

[①]* 也称"可指明"原因。但是有时我们面临的系统显然不在控制状态下，而我们又找不出原因。所以我们仍用 Deming 不过分乐观的词 —"特定"(special)。

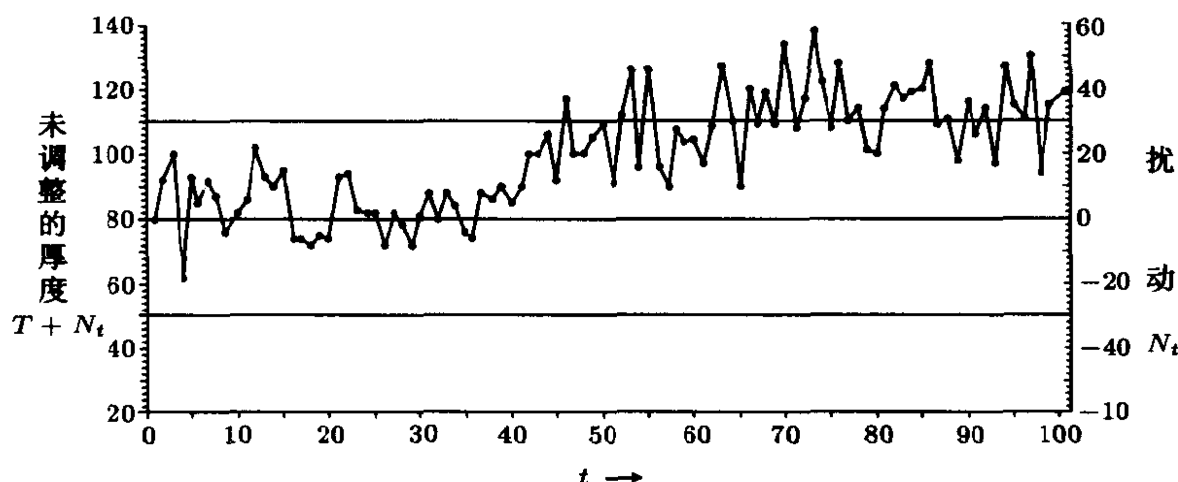


图 13.2 未进行调整的 100 个金属薄膜厚度值

差的偏为 $\epsilon_{t+1}=Y_{t+1}-T$, 在调整后将成为

$$\epsilon_{t+1}=gX_t+N_{t+1} \tag{13.1.1}$$

现在假设在 t 时刻你可以通过某种方式对 N_{t+1} 计算一个估计值(预报值) $\hat{N}_t(1)$, 并且该预报的误差为 $e_t(1)$, 从而

$$N_{t+1}=\hat{N}_t(1)+e_t(1) \tag{13.1.2}$$

那么，由 (13.1.1) 和 (13.1.2),

$$\epsilon_{t+1}=gX_t+\hat{N}_t(1)+e_t(1) \tag{13.1.3}$$

特别地，如果 X 可以调整以使在 t 时刻有

$$X_t=-\frac{1}{g}\hat{N}_t(1) \tag{13.1.4}$$

则对于调整后的过程有

$$\epsilon_{t+1}=e_t(1) \tag{13.1.5}$$

因此在进行调整后，过程对目标值的偏差 ϵ_{t+1} 就是对 N_{t+1} 的预报误差 $e_t(1)$, 而不是过程未调整时测得的偏差 N_{t+1} 。

如果我们用一个或多个已知的干扰输入因子(如周围温度)的测量来计算 N_{t+1} 的估计值 $\hat{N}_t(1)$, 这样我们就有了一个前馈控制的例子。如果 N_{t+1} 的估计值 $\hat{N}_t(1)$ 只是值接或间接地使用输出扰动的当前和过去值 $N_t, N_{t-1}, N_{t-2}, \cdots$, 那么, 方程 (13.1.4) 就定义了一个反馈控制系统。一个混合的前馈反馈控制系统同时使用两种数据. 为简单起见, 我们这里只考虑反馈控制的情形。

13.2 使用反馈控制的过程调整

经验调整的引入　(13.1.4) 中的估计值 $\hat{N}_t(1)$ 常常可以适当使用其过去值 $N_t, N_{t-1}, N_{t-2}, \cdots$ 的某种加权平均得出。特别地, 指数加权滑动平均 (EWMA) 在直观上具有吸引力, 因为它对最近发生的数据赋予了最大的权重。假设 $\hat{N}_t(1)$ 是一个 EWMA,

$$\hat{N}_t(1) = \lambda(N_t + \theta N_{t-1} + \theta^2 N_{t-2} + \cdots) \qquad (0 < \theta < 1) \qquad (13.2.1)$$

其中 θ 是平滑常数, $\lambda = 1 - \theta$ 。

我们首先考虑工业过程中很常见的一种情况, 即调整是在每一个观测到手时持续进行的。然后利用 (13.1.4) 式, t 时刻作出的调整 (沉淀速度的变化) 将由下式给出

$$X_t - X_{t-1} = -\frac{1}{g}[\hat{N}_t(1) - \hat{N}_{t-1}(1)] \qquad (13.2.2)$$

现在有了预报误差 $e_{t-1}(1) = N_t - \hat{N}_{t-1}(1)$, 可以把一个 EWMA 预报的修正公式写为

$$\hat{N}_t(1) - \hat{N}_{t-1}(1) = \lambda e_{t-1}(1) \qquad (13.2.3)$$

因此, 对于使用了补偿变量 X 来抵消噪声 $\{N_t\}$ 的 EWMA 反馈方案, 所要求的反馈量应满足

$$X_t - X_{t-1} = -\frac{\lambda}{g}e_{t-1}(1) = -\frac{\lambda}{g}\epsilon_t \qquad (13.2.4)$$

对于金属沉淀过程而言， $g=1.2, \lambda=0.2$, 及 $T=80$, 因此，调整方程为

$$X_t - X_{t-1} = -\frac{1}{6}\epsilon_t \tag{13.2.5}$$

13.2.1 反馈调整图

如图 13.3 所示，这类调整非常容易使用。图中是对前面给出的金属膜厚度一例的人工反馈调整图 (Box 和 Jenkins[48])．在使用时，操作人员记下厚度的上一个值，并从调整刻度上读出他或她当前应该对沉淀速度增加或减少的适当读数。例如，所记录的第一个厚度 80 正好是目标值，所以无须采取行动。第二个值 92 比目标值高 12 个单位，故 $\epsilon_2=12$, 左边刻度上相应的沉淀速度调整为 $X_2 - X_1 = -2$ ．所以操作员现在应使沉淀速度从其原来水平减少 2 个单位。

注意，这个图中所示的顺次记录的厚度值是调整以后实际发生的读数，当然从图上看不到隐含的扰动。在这个例子中，这个图在观测记录期间使均方误差减小了超过五倍。现在调整后厚度的标准差只有 $\sigma_\epsilon = 11$ ．请注意以下几点：

1. 该图并不比 Shewhart 图更难使用。

2. 尽管“直观”调整为 $-(1/g)\epsilon_t = -\frac{5}{6}\epsilon_t$(相应于 Deming 所

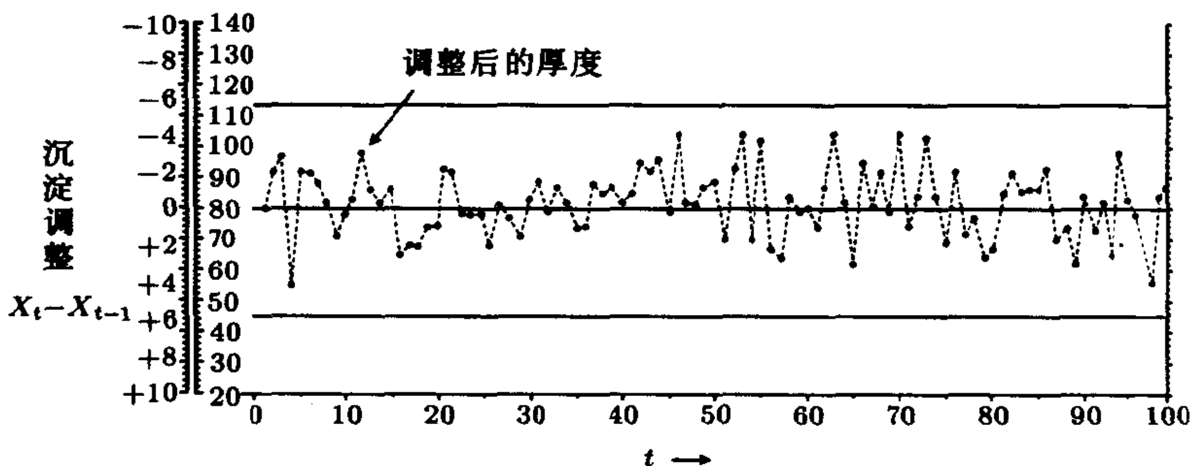


图 13.3 厚度的人工调整图，使操作员可以读到沉淀速度的适当改变量

称的“修补”(tinkering), 由式 (13.2.4) 给出的调整为 $-(\lambda/g)\epsilon = -\frac{1}{6}\epsilon_t$ 。因此它用的是相对于目标值偏离且打了折扣或减弱了的估计值 $\lambda\epsilon_t$, 以此来决定适当的调整，其折扣因子 λ 为 $1-\theta$, θ 是噪声 EWMA 估计的平滑常数。

3. 通过对 (13.2.4) 求和，我们看到 t 时刻的全部调整为

$$X_t = k_0 + k_I \sum_{i=1}^{t} \epsilon_i \tag{13.2.6}$$

其中 $k_0 = X_0, k_I = -\lambda/g$ 。因此这个调整过程依赖于调整误差的累积和，常数 k_I 确定了“直观”调整被折扣了多少。

4. 由前面的调整可知，这一调整等价于在时刻 t 用过去值的指数加权平均来估计下一个未调整的扰动 N_{t+1} 值，并利用这个估计值来作出适当的调整。在图 13.4 中对金属膜厚度一例作了详细说明。注意，在这一初步讨论中，我们并没有对该过程直接假设任何具体的最优属性。在某种程度上，可以从解释其显著稳健性的角度加以讨论，这些在后文将讨论。

总结如下：

1. 对于过程监视，我们所指的是比如 Shewhart 图或 Cuscore 图 (Box 和 Ramírez[57]) 的使用。这些方法用来不断地检查描述

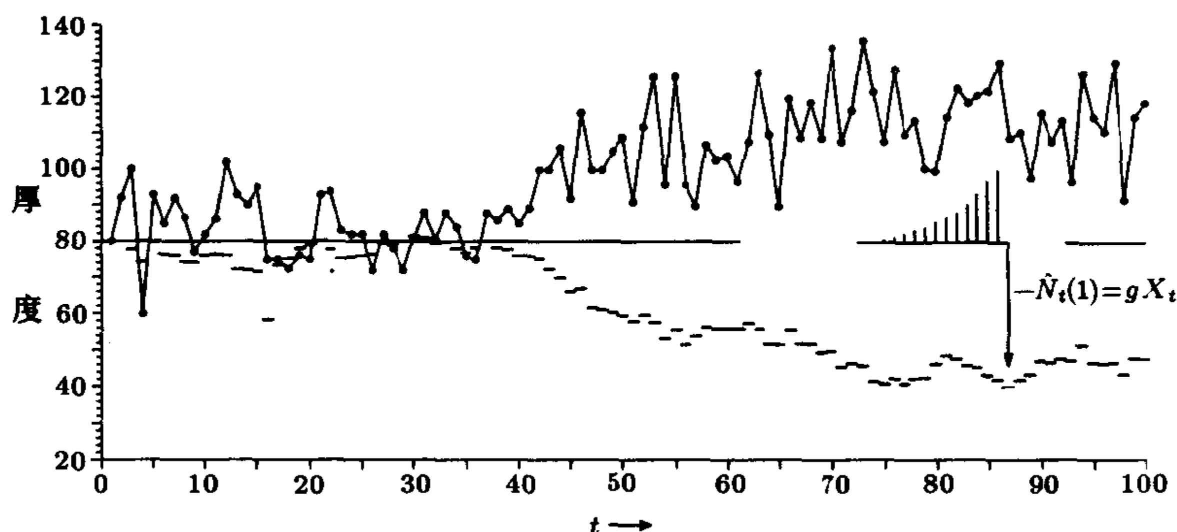


图 13.4 虚线所示为由图 13.3 人工调整得到的总调整量 $\hat{N}_t(1)=-gX_t$

系统理想稳定状态的模型：例如，在固定目标 T 附近的正态独立同分布 (NIID) 偏差。使用这些图可以引导消除由异常表现所指出的特殊原因。为了判断异常表现是否足以引起注意，可由一个类似于假设检验的过程来决定。其特征是通过概率形式加以描述(例如，一个点落在 Shewhart 图 3σ 限以外的概率)。

2. 对于过程调整，我们所指的是用反馈或前馈控制，或二者结合来保持过程尽可能地接近某一理想目标值。过程调整使用了统计估计 (预报) 系统，而不是假设检验系统。其特征是由例如输出的均方误差来描述的。对应于假设检验和估计的互补性，过程监视和过程调整也是互补的，而不是相互竞争的 (如见 Box[35])。我们将在本章的后面做更详细地讨论。

13.2.2 反馈闭环建模

图 13.5 所示为一个更一般的反馈控制系统。该过程受到扰动的影响，在没有补偿行为时，扰动会使输出的质量特征以量值 N_t 偏离其目标。因此，$\{N_t\}$ 就是这样的一个时间序列，它说明了当不施加控制时输出将会发生什么。事实上，可以操纵一个补偿变量 X_t (我们例子中的沉淀速度) 来尽可能地抵消这一扰动。X 的变化贯穿过程始终，并通过其动态作用对 t 时刻输出形成一个补偿量 $\mathcal{Y}_t$(仍由对目标值的偏离来度量)。从补偿 $\mathcal{Y}_t$ 中抵消扰动的不足部分，便是误差或对目标值的偏离 $\epsilon_t = Y_t - T$，它等于 $\epsilon_t = N_t + \mathcal{Y}_t$。控制器就是这样的一些装置 (自动的或人工的)，它使控制方程 $X_t = f(\epsilon_t, \epsilon_{t-1}, \cdots)$ 生效，而控制方程根据现在和过去误差来调整输出。

一种在工业过程中使用了多年的装置是三项控制器。这种控制器通常是自动操作的。使用连续而不是离散的测量和调整。如果 ϵ_t 是 t 时刻输出的误差，则控制行为可由 ϵ 自身，或 ϵ 关于时间的积分或微分的比例而作出。一个三项控制器使用的是这些控制行为模型的线性组合，因此，如果 X_t 表示 t 时刻操纵变量的

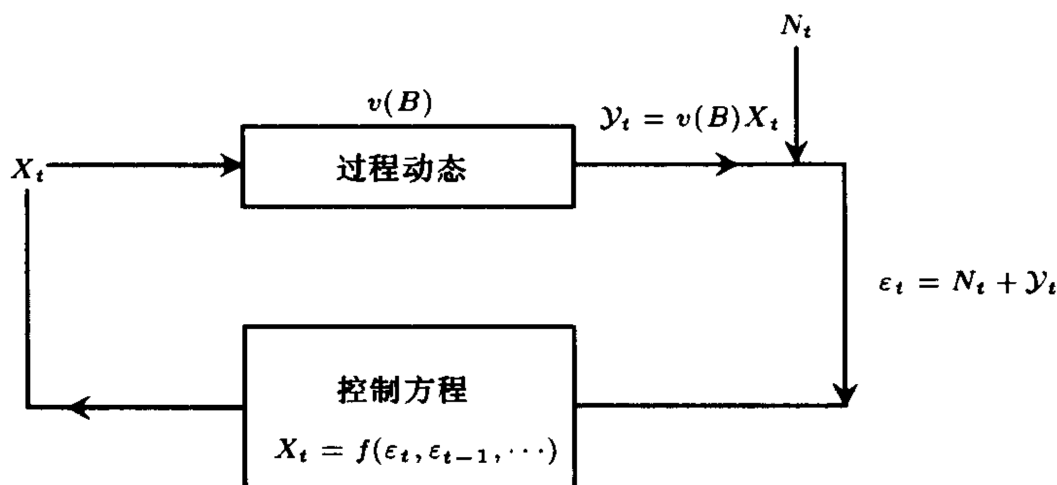


图 13.5 反馈控制环

水平，那么控制方程形式为

$$X_t = k_0 + k_D \frac{d\epsilon_t}{dt} + k_P \epsilon_t + k_I \int \epsilon_t dt \qquad (13.2.7)$$

这里 k_D, k_P 和 k_I 都是常数。通常只使用三项模型中的一项或两项控制行为。具体地，若只有 k_I 非零 $(k_d = 0, K_P = 0)$，我们得到积分控制。若只有 k_I 和 k_P 非零 $(k_D = 0)$，我们得到比例 - 积分 (PI) 控制。

注意，在我们刚刚讨论的例子中，任何调整结果在一个时间间隔内对输出完全生效，过程的动态特性表示为 $\mathcal{Y}_t = gX_{t-1} = gBX_t$ 。(13.2.6) 中的控制方程 $X_t = k_0 + k_I \sum_{i=1}^{t} \epsilon_i$ 就是这个控制工程中积分控制的离散形式。

一般地，(13.2.7) 的离散形式为

$$X_t = k_0 + k_D \nabla \epsilon_t + k_P \epsilon_t + k_I \sum_{i=1}^{t} \epsilon_i$$

或者用调整的形式给出

$$\begin{aligned} x_t = X_t - X_{t-1} &= k_D \nabla^2 \epsilon_t + k_P \nabla \epsilon_t + k_I \epsilon_t \\ &= c_1 \epsilon_t + c_2 \epsilon_{l-1} + c_3 \epsilon_{t-2} \end{aligned}$$

其中 c_1, c_2 和 c_3 是适当常数。不难料想，这种控制方法具有相当高的实用价值。

13.2.3 扰动和动态特性的简单模型

到目前为止，我们已在完全经验的基础上介绍了一个简单的反馈控制系统。这种系统的效率取决于扰动和过程的动态特征。从理论的角度来看，我们可以对噪声和动态特征考虑非常一般的模型，然后根据某些准则找到使系统最优的控制方程。但是这些模型的实际效率往往取决于：模型及“最优”准则是否具有广泛的科学含义，模型对理想目标的可能偏差是否稳健。一开始从单纯常识角度讨论控制过程时，我们就将此牢记于心，今后在为扰动和过程动态选择模型时，我们将一如既往。

用方差图来刻划适当的扰动模型　一种有助于刻划过程扰动的工具是标准化的方差图，它所度量的是相隔 m 步观测之差的方差与相隔 1 步观测之差的方差二者之比：

$$G_m = \frac{\mathrm{var}[N_{t+m} - N_t]}{\mathrm{var}[N_{t+1} - N_t]} \equiv \frac{V_m}{V_1} \tag{13.2.8}$$

对于平稳过程， G_m 是自相关函数的一个简单函数。事实上，$G_m = (1-\rho_m)/(1-\rho_1)$ 。但是，方差图不仅可以用于刻划平稳特征，还可以用于刻划非平稳特征。图 13.6 所示为从目标值开始的 100 个观测值的实现，它们是由如下各模型生成：(a) 白噪声过程，(b) 一阶自回归过程，及 (c,d,e,f)IMA(0,1,1) 过程，分别具有 $\lambda = 0.1, 0.2, 0.3, 0.4$ 。图中还给出了这些时间序列模型相应的标准化方差图。

在某种想象的世界里，我们可以一劳永逸地确定一台机器的控制，并且将一组指令给予一个永远警觉、永不遗忘的操作员，由此将会得到完全稳定的过程。在这种情况下，扰动可以由白噪声来表示，它相应的标准化方差图 G_m 将独立于 m 并等于 1. 但是在现实世界里，生产中放任自由的机器会逐渐失去调节并被损

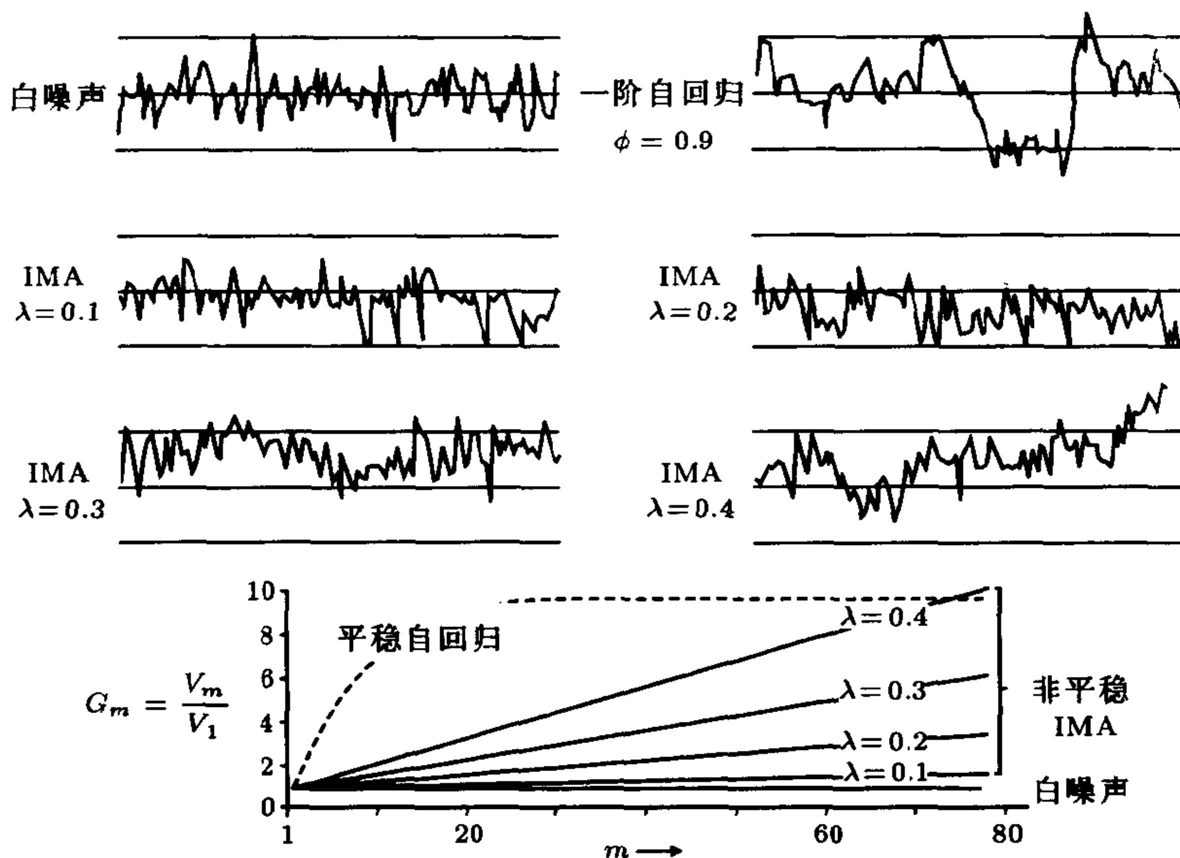


图 13.6 白噪声，自回归和 IMA(0,1,1) 时间序列的实现及理论方差图

坏，放任自由的人们会逐渐忘掉指令和联络。因此对于一个未控制的扰动，可以料想方差图以某种形式单调上升。对于一个线性平稳模型我们得不到这样的方差图，因为尽管 G_m 在开始时随着 m 而增加，但它总会达到某一渐近线．这一过程可以非常迅速的发生，甚至在顺序观测高度正相关时也是如此。图中所示的一阶平稳自相关时间序列模型 $N_t = 0.9N_{t-1} + a_t$ 的方差图说明了这一点。在此例中，尽管对目标偏差顺序值 N_t 的自相关系数为 0.9,G_m 仅在 20 步之后就已处于其渐近值的 5% 之内了。这意味着，比如，对于该模型生成的观测值， 100 步之后与 20 步之后就没有差别了。

对于熵持续增长的无控制系统，其特征可以通过把扰动考虑为包含两个部分的模型来近似：一个瞬时部分 b_t 和一个非瞬时部分 w_t:

$$N_t = b_t + w_t \tag{13.2.9}$$

瞬时部分 b_t 仅与第 t 个观测值相联系，且假设它独立于其它时刻的观测。一般引起 b_t 的原因是测量和抽样误差。我们把这一瞬时性部分表示为随机抽取一个零均值方差为 σ_b^2 的分布。

“粘性新息”模型　非瞬时部分 w_t 所表示的是新息，它不时地进入系统中，并粘着在那里。这些“粘性”新息可能来自多种原因，比如磨损，腐蚀，人为失误等等。例如，汽车轮胎碰到尖利的石头后，轮胎花纹就受到些许损伤；一个主动轴表面由于腐蚀而出现的小坑会保留在那里；在一个医院里，取血压的标准过程中某个细节被遗忘后就长久地被遗漏或改变。正是这些非瞬时或粘性新息构成了我们希望将其消除的非需要信号。每个系统都受制于这种影响。如果不采取行动来制止它们，将持续地使熵提高。 Barnard[21] 给出了粘性新息的模型，其方差图随着 m 线性上升．这种模型的一个特例是 IMA(0,1,1) 模型

$$N_t - N_{t-1} = a_t - \theta a_{t-1} \tag{13.2.10}$$

它可以用于近似该模型。因为对于这一模型来说，具有平滑参数 θ 方程 (13.2.1) 的 EWMA 提供了一个最小均方误差 (MMSE) 预报，预报误差 $e_{t-1}(1) = a_t$ 。所以具有 $k_l = -\lambda/g$ 的 (13.2.6) 的离散“积分控制器”形成了 $\epsilon = a_t$ 的 MMSE 控制器。这是一般 MMSE 线性反馈控制方案的一个特例，我们以后将更正式地讨论。

动态特性　在 (13.2.6) 积分方案的讨论中，我们假设对系统输入的任一变化都会在一个时间间隔内对输出完全生效，因此响应效果 $\mathcal{Y}_t$ 的假设动态方程为

$$\mathcal{Y}_t = gBX_{t+} \tag{13.2.11}$$

一个更一般的假设是系统可以描述为一阶差分方程

$$(1 + \xi\nabla)\mathcal{Y}_t = gBX_{t+} \tag{13.2.12}$$

[如见 (10.3.6)] 或等价地,

$$(1-\delta B)\mathcal{Y}_t=(1-\delta)gBX_{t+} \qquad (-1<\delta<1) \qquad (13.2.13)$$

其中 $\xi=\delta/(1-\delta)$ 或等价地, $\delta=\xi/(1+\xi)$ 。在这种情形下, 在 $t+1$ 时刻 [参见 (13.1.1)] 对目标的偏离经调整后为

$$\epsilon_{t+1}=\mathcal{Y}_{t+1}+N_{t+1}$$

从而

$$\epsilon_{t+1}=\frac{(1-\delta)g}{1-\delta B}X_{t+}+\hat{N}_t(1)+e_t(1)$$

其中 $\hat{N}_t(1)$ 是在时刻 t 作出的 N_t 的某种预报, 预报误差为 $e_t(1)$ 。如果我们使用调整方程

$$X_{t+1}-X_t=x_t=-\frac{1-\delta B}{(1-\delta)g}[\hat{N}_t(1)-\hat{N}_{t-1}(1)]$$

可知, 对目标的偏离 ϵ_{t+1} 等于预报误差 $e_t(1)$ 。这样我们又可以用预报 N_{t+1} 时的误差来代替偏差 N_{t+1} 自身。具体地, 如果 $\hat{N}_t(1)$ 是平滑参数为 θ 的一个 EWMA 预报, 并且 $\lambda=1-\theta$, 那么, 利用 (13.2.3), 有

$$x_t=(1-B)X_t=-\frac{\lambda(1-\delta B)}{g(1-\delta)}\epsilon_t=-\frac{\lambda(1-\delta)+\delta\nabla}{g(1-\delta)}\epsilon_t \quad (13.2.14)$$

最后, 若 N_t 可以表示为一个 IMA(0,1,1) 过程, 参数为 θ, 则 $\epsilon_t=a_t$, 且这一调整会得到 MMSE 控制。对 (13.2.14) 求和, 我们得到

$$X_t=k_0+k_P\epsilon_t+k_I\sum_{i=1}^{t}\epsilon_i \qquad (13.2.15)$$

其中

$$k_P=-\frac{\lambda}{g}\xi \quad 和 \quad k_I=-\frac{\lambda}{g}$$

控制方程 (13.2.15) 得到了与前面提到的连续比例积分 (PI) 控制的相似形式，因此被称为 (离散)PI 控制。

注意，相应于离散积分控制和比例积分控制的调整方程尽管存在着重要差异，但它们都是极其简单和直观的。对于离散积分控制

$$x_t = c_1\epsilon_t \qquad (\text{其中} c_1 = k_I)$$

对于比例积分控制

$$x_t = c_1\epsilon_t + c_2\epsilon_{t-1} \qquad (\text{其中} c_1 = k_I + k_P \text{ 和} c_2 = -k_P)$$

这样它们分别使调整量 x_t 线性依赖于上一期和上两期误差。

13.2.4 一般最小均方误差反馈控制方案

前面提到，对于第 4 章和第 10 章所讨论的更一般的随机线性动态模型而言，推导理论上的最小均方误差控制方案并不困难。假设可控制输入变量为 X_t，其调整序列的响应由动态传递函数关系 (10.2.3) 表示，写作

$$\mathcal{Y}_t = L_1^{-1}(B)L_2(B)B^{f+1}X_{t+}$$

其中 $L_1(B)$ 和 $L_2(B)$ 是 B 的多项式。这一关系允许响应中有 P 期停滞时间。假设噪声或过程扰动 $\{N_t\}$ 可用线性随机 ARIMA 过程表示，定义为

$$N_t = \varphi^{-1}(B)\theta(B)a_t = (1 + \sum_{i=1}^{\infty}\psi_i B^i)a_t$$

其中 a_t 是白噪声过程。那么，输出误差 $\epsilon_{t+f+1} = Y_{t+f+1} - T$，在 $t + f + 1$ 时刻可以写为

$$\epsilon_{t+f+1} = \mathcal{Y}_{t+f+1} + N_{t+f+1} = L_1^{-1}(B)L_2(B)X_{t+} + N_{t+f+1}$$

现在我们可以写成 $N_{t+f+1}=\hat{N}_t(f+1)+e_t(f+1)$, 其中 $\hat{N}_t(f+1)$ 是时刻 t 对 N_{t+f+1} 的预报, $e_t(f+1)$ 是提前 $f+1$ 期外推预报误差。噪声 N_{t+f+1} 在时刻 t 是未知的, 但它的最小均方误差预报 $\hat{N}_t(f+1)$ 可以由误差序列 $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \cdots$ 推出, 这些是已观测到的。因此, 控制方程 $X_{t+}=-L_1(B)L_2^{-1}(B)N_t(f+1)$ 将会在时刻 $t+f+1$ 对输出产生一个作用来抵消噪声的 $f+1$ 期外推预报, 从而输出的误差就是 $\epsilon_{t+f+1}=e_t(f+1)$, 即预报误差。为了用误差序列 ϵ_t 来表示控制方程, 我们可以写为

$$\epsilon_t = e_{t-f-1}(f+1) = a_t + \psi_1 a_{t-1} + \cdots + \psi_f a_{t-f} = L_4(B)a_t$$

及

$$\hat{N}_t(f+1) = \psi_{f+1}a_t + \psi_{f+2}a_{t-1} + \cdots = L_3(B)\epsilon_t$$

其中算子模型 $L_3(B)$ 和 $L_4(B)$ 由噪声过程模型 $N_t=\varphi^{-1}(B)\theta(B)a_t=\phi(B)a_t$ 所确定。因此我们有

$$\hat{N}_t(f+1) = L_3(B)L_4^{-1}(B)\epsilon_t$$

从而最小均方误差 (MMSE) 反馈控制方案为

$$X_{t+} = -\frac{L_1(B)L_3(B)}{L_2(B)L_4(B)}\epsilon_t \qquad (13.2.16)$$

作为另一种方便的替换方式, 我们可以用 t 时刻作出的调整 $x_t=X_{t+}-X_{t-1+}$ 来定义控制行动

$$x_t = -\frac{L_1(B)L_3(B)(1-B)}{L_2(B)L_4(B)}\epsilon_t$$

例: 有停滞时间的模型　具体地说, 上面使用的一般动态模型允许"停滞时间" — 即对调整的响应有纯延迟. 为了举例说明方程 (13.2.16) 的应用, 考虑一个一阶系统, 影响它的纯延迟在第 f 和第 $f+1$ 和单位间隔之间, 从而

$$(1-\delta B)\mathcal{Y}_t = g(1-\delta)[(1-\nu)+\nu B]B^f X_{t-1} \qquad (13.2.17)$$

将它和方程 (13.2.10) 的扰动模型 IMA(0,1,1) 合并起来，我们可以用上面的一般推导来得到最小均方误差控制方案。用一般模型来表示，我们有 $L_2(B)/L_1(B) = g(1-\delta)(1-\nu\nabla)/(1-\delta B)$,IMA 噪声模型得出 $\hat{N}_t(f+1) - \hat{N}_{t-1}(f+1) = \lambda a_t$，从而 $L_3(B) = \lambda/(1-B)$，还有

$$e_{t-f-1}(f+1) = [1 + \lambda(B + B^2 + \cdots + B^f)]a_t \equiv L_4(B)a_t$$

因此，对于调整 x_t 我们有关系式 $L_2(B)L_4(B)x_t = -L_1(B)L_3(B)(1-B)\epsilon_t$，并得到 MMSE 控制方程为

$$(1-\nu\nabla)[1 + \lambda(B + B^2 + \cdots + B^f)]x_t = -\frac{\lambda}{g(1-\delta)}(1-\delta B)\epsilon_t$$

因此这一最优控制方案不是 PI 而是形如

$$x_t = c_1 x_{t-1} + c_2 x_{t-2} + \cdots + c_f x_{t-f-1} + c(\epsilon_t - \delta\epsilon_{t-1}) \quad (13.2.18)$$

这里 $c = -\lambda/[g(1-\delta)] = k_I + k_P$ 。

Fearn 和 Maris[90] 给出了一个有趣的例子，所描述的是一个 MMSE 方案，被应用于在面粉厂中控制向做面包的面粉中添加面筋，这里的目的是使面粉中的蛋白质含量尽可能接近目标。详细的研究过程表明，对于该过程的恰当近似为 $\delta = 0, \nu = 0, f = 1$，及 $\lambda = 0.25(\theta = 0.75)$ 。因此调整方程为

$$x_t = -0.25x_{t-1} - \frac{0.25}{g}\epsilon_t \quad (13.2.19)$$

经过对该方案的大量检验，作者指出它在很大范围的生产条件下都运作良好，并且对于参数的适度变化是稳健的。

这个面粉生产的例子并未得出 PI 方案，然而应注意，调整方程可写为 $x_t = -(1+\lambda B)^{-1}(\lambda/g)\epsilon_t = -(1-\lambda B+\lambda^2 B^2-\cdots)(\lambda/g)\epsilon_t$ 。对于相当小的 $\lambda = 0.25$，如果我们截掉第一项以后的展式，则可

以得到 PI 方案 $x_t = c_1\epsilon_t + c_2\epsilon_{t-1}$, 其中 $c_1 = -\lambda/g, c_2 = \lambda^2/g$ 。实际中这个 PI 方案的特征几乎等同于 (13.2.19) 。更一般地，我们可以发现 PI 方案对于某些简单模型来说，除了它们得到 MMSE 方案以外，还具有别的重要性。因此，我们下面考虑如何利用简单的反馈控制图来使 PI 方案生效。

13.2.5 离散比例积分方案的人工调整

离散 PI 方案 (13.2.15) 的控制方程 $x_t = X_t - X_{t-1}$ 还可以写为

$$x_t = -G(1 + P\nabla)\epsilon_t \tag{13.2.20}$$

其中

$$-G = k_I \qquad \text{和} \qquad P = \frac{k_P}{k_I} \tag{13.2.21}$$

或等价地 $k_I = -G$ 和 $k_P = -PG$, 对于纯积分控制有 $P = 0$ 。在随机模型和动态模型分别由 (13.2.10) 和 (13.2.12) 定义的特例中，PI 控制方程 (13.2.15) 在 $G = \lambda/g$ 和 $P = \xi$ 时得到 MMSE 。

式 (13.2.20) 说明了我们能如何制造一个人工调整图来使 PI 控制生效。对于金属膜厚度一例，我们已在图 13.3 中详细说明了这种图的使用。为了进一步说明，我们采用 Box,Hunter 和 Hunter[41] 讨论的一个例子。在一个染色过程中，感兴趣的指标是色彩指标。对于理想目标值 $T = 9$ 的偏离 ϵ_t 可以通过染料添加速度 X 来补偿。对于这个例子，色彩指标中的扰动可以由一个 IMA(0,1,1) 模型来近似，其中 $\lambda = 0.3$, 染料添加速度 X 每 1 单位的变化将最终引起色彩指标 0.06 单位的变化，因此 $g = 0.06$ 。

首先假设 ξ 为零，从而动态模型就是 $Y_t = gBX_t$, 这意味着输入 X 的变化在一个时间间隔内对输出完全生效。那么

$$-G = k_I = -\frac{\lambda}{g} = -\frac{0.30}{0.06} = -5 \qquad \text{和 } k_P = 0 \tag{13.2.22}$$

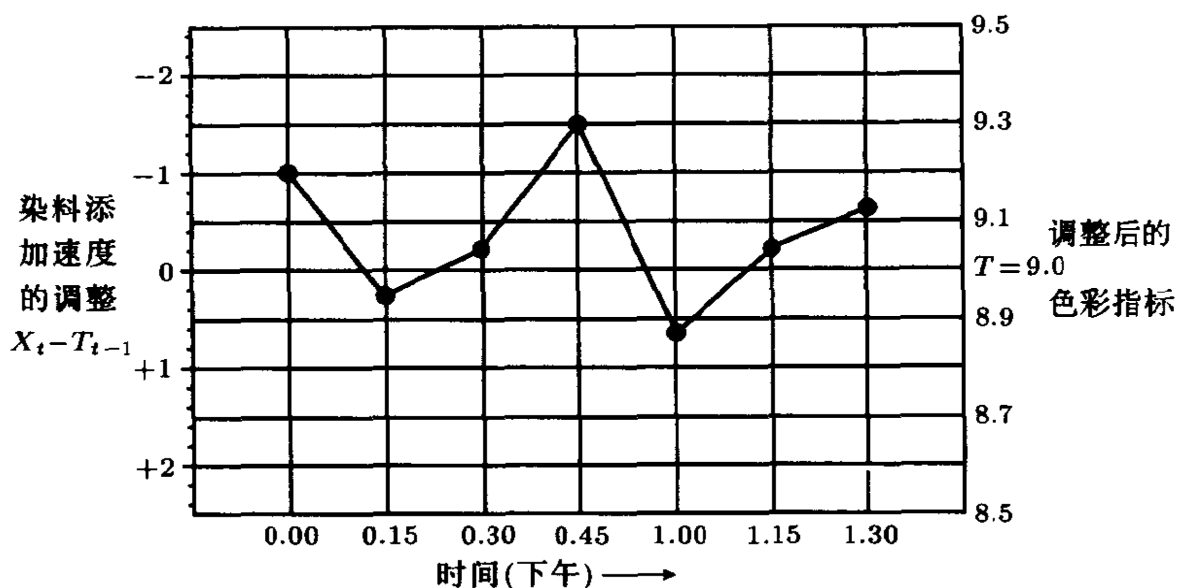


图 13.7　离散积分控制的人工调整图

MMSE 积分反馈控制方案为

$$X_t = k_0 - G\sum_{i=1}^{t}\epsilon_t = k_0 - 5\sum_{i=1}^{t}\epsilon_i \tag{13.2.23}$$

相应 t 时刻的调整为

$$x_t = -G\epsilon_t = -5\epsilon_t \tag{13.2.24}$$

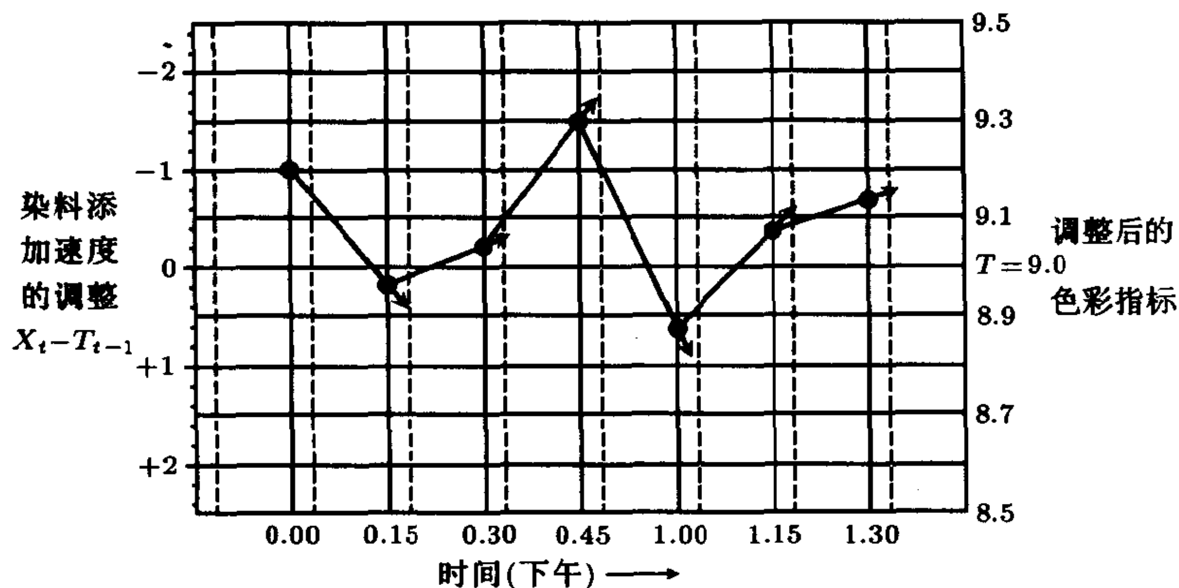


图 13.8　实施离散积分加上比例控制的人工调整图

从图 13.7 的人工调整图可以读到调整行动，其刻度是色彩指标 1 单位的偏离对应于染料添加速度 $-G=-5$ 单位的调整。每一个观测之后都采取行动。记下色彩指标的值 (由实点所示)，并从右边的刻度上读出染料添加速度所需的调整。因此，在下午 1:30 的图上，色彩指标为 9.14，它需要将染料添加速度降低 0.7.

现在考虑这样的情形：也许是由于染料的不完全混合，该过程受惯性的制约，这可以用形如 (13.2.13) 的一阶动态系统来近似，$\delta=0.2$，因此，$\xi=\delta/(1-\delta)=0.25$ 。这样，与前面一样，$G=0.3/0.06=5$，而现在 $P=\xi=0.25$ 。所以适当的 MMSE 控制方程 (13.2.15) 需要比例积分行动，使得

$$X_t=k_0-1.25\epsilon_t-5\sum_{i=1}^{t}\epsilon_i \qquad (13.2.25)$$

相应的调整方程为

$$x_t=-5(1+0.25\nabla)\epsilon_t \qquad (13.2.26)$$

为了人为地使其生效，可以使用图 13.8 中的图形，在每个采样间隔的 $P=k_P/k_I=0.25$ 处设置了竖直的虚线。在每一步中，操作员将上两点的连线外推至下一条虚线，读出适当的调整量。所以在图中，上两个读数在 1:15 和 1:30 分别为 9.06 和 9.14. 虚拟值 9.16 需要染料添加速度降低 0.8 个单位。并不需要很精确，由眼睛外推出来的线就够好的了。在我们开发 PI 图的其它应用时，有时会用到 P 的负值的方案，这时需要在上面两点之间内插而不是外推。

范围调整　到现在为止所讨论的反馈方案要求在每一种可能都采取某种行动 — 本例中是每 15 分钟。在实际中，若使用图 13.9 所示的按范围调整，并不会有多少损失。在原图上把行动刻度分成几个范围，这样就可很容易地构造出一个图。一个观测值

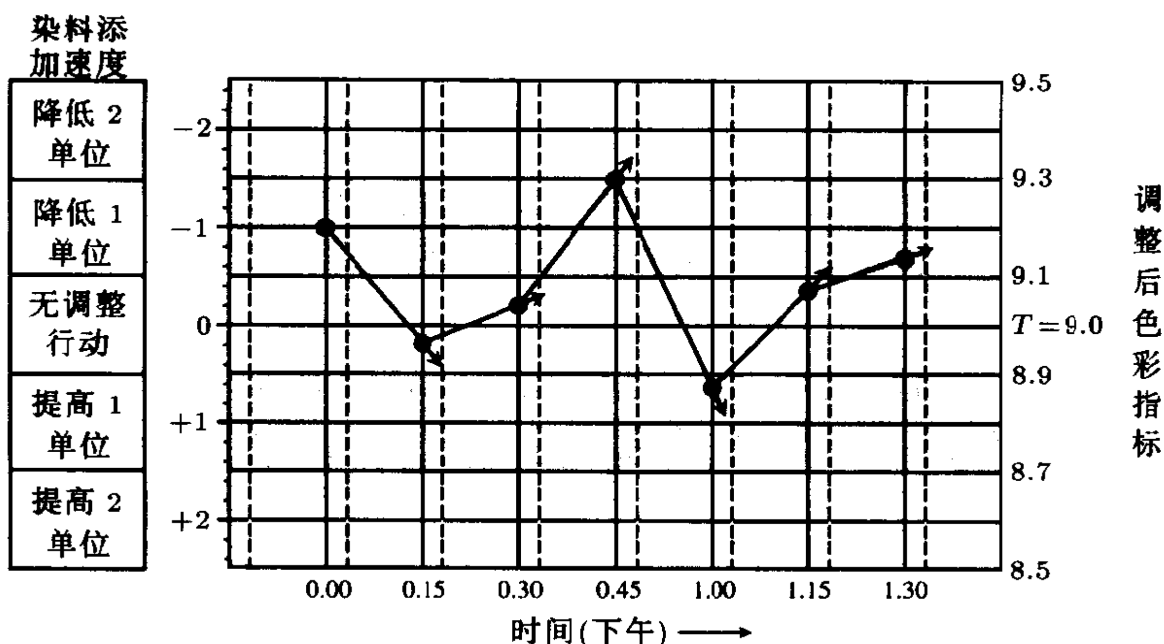


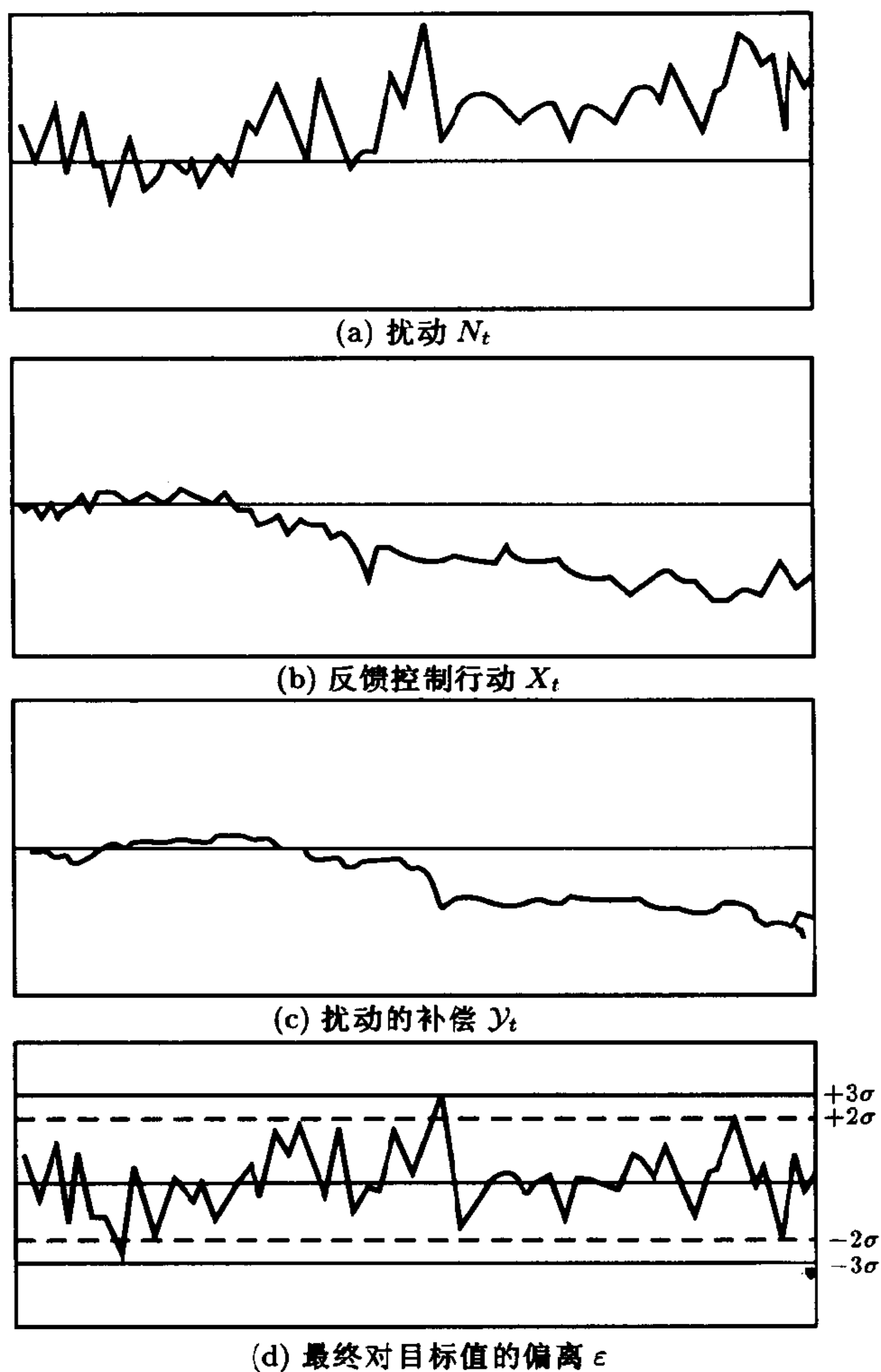
图 13.9 比例积分控制的范围调整

落入某个范围时所作的调整近似于原图上该范围的中点。图 13.9 给出了一个范围图，其中可能的行动仅限于 −2-,−1-,0-,1- 和 2- 个单位的催化剂方程式的变化。由于使用范围调整方案可能导致均方误差的增加 (通常很小)，但由于只需要对数目很少的标准调整进行工作，这将带来极大的便利。方便的范围宽度大约是一个标准差 σ_ϵ 或略少. Box 和 Jenkins[48,13.1 节] 提供了使用这种图的修正，在那里考虑了在调整量 x_t 中误差的影响. 注意，使用所有这些人工调整都不需要计算 — 它们简单而且完全图形化。

13.2.6 监视和调整的互补作用

有时人们抱怨反馈控制掩盖了被补偿扰动的特征，不然的话就可以将其消除掉. 但是如果与适当的监视结合起来，这种情形就不会发生了。调整方案和监视方案是互补的，应该协调使用。图 13.10 说明了这一点. 该图表现出了模拟反馈方案的特征，其中扰动是一个 IMA(0,1,1) 过程， $\lambda = 0.2$, 过程的动态特性由一个一阶系统 (13.2.13) 表示， $\delta = 0.5, g = 1.0$. 假设系统由 PI 控制器所控制，通过计算

$$-X_t = \text{常数} + 0.20\epsilon + 0.20\sum_{i=1}^{t}\epsilon_i \qquad (13.2.27)$$



(a) 扰动 N_t

(b) 反馈控制行动 X_t

(c) 扰动的补偿 $\mathcal{Y}_t$

(d) 最终对目标值的偏离 ε

图 13.10

对于这些列出的参数值，得到 MMSE 。尽管通常并不这样做，但图 13.10(b) 中的控制行动 X_t 和图 13.10(b) 中的目标偏离 $\{\epsilon_t\}$ 是可以画出来的 (或者更好地，显示在一个过程计算机的屏幕上) 。假设动态特性已知，图 13.10(c) 所示的精确补偿 $\mathcal{Y}_t$ 可以

计算，从而可以重现出图 13.10(a) 的原始扰动 N_t 。

考察这些监视可以引出一般和具体原因的推广概念。扰动和动态系统一起定义了一般原因系统，在控制器的设计中已经考虑了。但是管理行动可以改变系统以及控制的适当形式。例如，假设在系统的运行中发现图 13.10(b) 中所示的反馈控制行动 X_t 的态势反映出原材料不纯净。如果检查出的这个关系是一个因果关系，则可以决定管理行动来改变控制系统，或者在原材料到达生产过程之前就除去其不纯净。或者如果这不可能或代价太大，则测量它们并通过适当的前馈控制来补偿。

另外，具体原因所引起对隐含系统模型的瞬时偏离，可能是由于控制器的错误运行或操作员的过失所引起的，它能在残差序列 $\{\epsilon_t\}$ 中显示出来，从而引发纠正行动。为了对此详细说明，对图 13.10(a) 中扰动 N_t 的第 30 个值加了一个 $3\sigma_a$ 的偏差。这个扰动在经过反馈控制后，该异常值在图 13.10(d) 中所示对目标偏离 ϵ_t 的记录中清晰可见，这被画成了一个 Shewart 图。控制限可由设计控制器所用的模型或稳定运行中 ϵ_t 的记录直接计算。此外，正如后面 13.5 节所指出，可以使用更细致的检查来检测系统参数的可能变化。

假设模型正确，此例中残差 ϵ_t 是一个白噪声序列。但是对于那些不是 MMSE 或允许停滞时间的控制方案，序列 $\{\epsilon_t\}$ 一般是自相关的。我们可以采用这样一种办法，即对 $\{\epsilon_t\}$ 适当地滤波来生成一个序列，它在给定的假设模型下是白噪声，然后对这一序列可以使用适当的检验。

13.3 MMSE 控制有时所需的过度调整

使用积分控制和比例积分控制的一个原因是，对于那些近似于现实且可能是最简单的扰动模型 (式 (13.2.10) 和动态模型 (式 (13.2.12) 和 (13.2.13))，这些形式的反馈控制可以得出最小均方误

差[1] 遗憾的是，MMSE 有时需要对补偿变量 X 进行不可接受的大量操纵。为了举例说明，还是考虑这样的情形：在适当的近似下，扰动模型是 (13.2.10) 的 IMA 模型，参数为 θ, 动态模型是一阶差分模型 (13.2.13), 参数为 δ 和 g 。那么，MMSE 反馈控制调整方案可以写作 (见 (13.2.14))

$$x_t = -\frac{\lambda}{g}\frac{1-\delta B}{1-\delta}\epsilon_t \tag{13.3.1}$$

其中 $\lambda = 1-\theta, \epsilon_t = a_t$ 。若 δ 小到可以忽略，MMSE 控制可以得到 $x_t = -(\lambda/g)\epsilon_t$, 然后我们写出下式：

$$\sigma_x^2 = \mathrm{var}[x_t] = \frac{\lambda^2}{g^2}\sigma_a^2 = k \tag{13.3.2}$$

但是若 δ 不可忽略，$\sigma_x^2 = k[(1+\delta^2)/(1-\delta)^2]$ 。因此若 δ 接近其上界 1 时，σ_x^2 可以非常大。例如，$\delta = 0.9$(因此在第一个间隔中阶跃输入引起的最终变化只发生了 1/10),$\sigma_x^2 = 181k$ 。事实上，随着 δ 接近于 1, 控制行动

$$x_t = -\frac{\lambda}{g(1-\delta)}(\epsilon_t - \delta\epsilon_{t-1}) \tag{13.3.3}$$

越来越表现出交替特征[2] 。t 时刻作出的调整复原了 $t-1$ 时刻所作调整的极大部分。这种交替变动调整的原因可以这样来理解：$\delta = 0.9$ 时，MMSE 控制的人工调整图中常数 $P = \xi = 9$, 它需要

---

[1] 这一理论形式来源于能得出 MMSE 的离散控制器，它并不是唯一的．例如，一个给出 MMSE 的 PI 控制器不仅可由动态模型 (13.213) 和 IMS(0,1,1) 噪声模型 (13.2.10) 得到，还可以由模型 $\mathcal{Y}_t = gBX_t$ 和 $N_t = (1-\theta_1 B - \theta_2 B^2)a_t$ 得到．

[2] δ=0.9 这个值对应于系统在 9 个采样间隔上的时间常数．这个值的出现立即提出这样的问题：采样间隔是否取得太短；事实上，过程的惯性如此之大，以致监视频率小一些几乎不会有损失．在这里 (见附录 A13.2) 采样间隔的选择问题必须取决于影响系统的噪声性质．由于噪声的特征通常反映了系统的惯性，所以在很多情形下可以得出的结论是：采样间隔应予以延长．

把联结 ϵ_{t-1} 和 ϵ_t 的直线外推 9 个采样区间. 实际中可以使用有约束的方案, 它以输出 MSE 相当小的提高为代价, 而要求的补偿操作则少得多.

13.3.1 有约束控制

当调整量 x_t 成为一个平稳时间序列时, 这时的有约束方案可以通过寻找下式的无约束最小值而得到

$$\sigma_\epsilon^2 + \alpha\sigma_x^2 \tag{13.3.4}$$

其中 α 可视为一个未定乘数, 它分配了 ϵ_t 和 x_t 变动的相对二次代价. 这样的方案可被称为有约束的 MMSE 方案或 CMMSE 方案. 具体地, 我们已经看到, 对于一个 IMA(0,1,1) 扰动和一阶动态无约束 MMSE 方案需要的调整为

$$x_t = -\frac{\lambda}{g}(1+\xi\nabla)\epsilon_t = -\frac{\lambda(1-\delta B)}{g(1-\delta)}\epsilon_t \tag{13.3.5}$$

附录 A13.1(式 (A13.1.27)) 证明了相应的 CMMSE 形式为

$$x_t = [k_I + (1-\lambda)k_0]x_{t-1} - (1-\lambda)k_I x_{t-2} - \frac{\lambda(1-k_0)(1-\delta B)}{g(1-\delta)}\epsilon_t \tag{13.3.6}$$

其中 k_0 和 k_I 是参数 g, λ, δ 和 α 相当复杂的函数. 附录 A13.1 还给出了应用这种控制的表.

为了详细说明, 假设 $\lambda = 0.6, \delta = 0.5, g = 1$, 那么, 最优无约束 MMSE 方案是

$$x_t = -1.2(1-0.5B)\epsilon_t \tag{13.3.7}$$

其中 $\sigma_x^2 = 1.80\sigma_a^2, \sigma_\epsilon^2 = \sigma_a^2$. 假设调整 x_t 中这么大的变动在过程操作中有困难, 需要将它降低以使 σ_x^2 约为 $0.50\sigma_a^2$. 使用表 A13.2 表明, 这可以通过如下方案来实现

$$x_t = 0.32x_{t-1} - 0.06x_{t-2} - (0.57\times 1.2)(1-0.5B)\epsilon_t \tag{13.3.8}$$

它使 σ_x^2 减少到 $0.47\sigma_a^2, \sigma_\epsilon^2 = 1.07\sigma_a^2$。这样，所得的 σ_a^2 减少到几乎是原来的 1/4, 而输出的方差仅提高了 7%。这样的最优有约束方案是极有吸引力的，因为它们以 σ_ϵ^2 很小的增大而得到 σ_x^2 的大大降低。可参见，如 Whittle[198],Wilson[203],[204],MacGregor[141], Box 和 Jenkins[48],Harris,MacGregor 和 Wright[104],Aström 和 Wittenmark[16], Rivera,Morari 和 Skogestad[169],Bergh 和 MacGregor[30]。不幸的是，这些方案可能会弄得很复杂。

但是在实际中，完全"最优"在某种程度上只是幻想，因为假想条件是不能实现的。人们总会找出一种几乎可与 CMMSE 控制相媲美的有约束控制，它往可以通过一个适当调节过的 PI 控制来得到. 这样的控制器其优越性是简单，尤其容易适用于人工控制。下面的例子说明了适当调整后的 PI 控制器如何可以与最优约束方案做得一样好，仅以输出方差 σ_ϵ^2 的些微提高而得到调整方差 σ_x^2 的大大降低。

为了举例说明，再考虑如下情形：过程扰动由 (13.2.10) 的 IMA(0,1,1) 过程表示，过程动态由一阶系统 (13.2.13) 表示，假设 $\lambda = 0.4, \sigma_a^2 = 1, g = 0.4, \delta = 0.5$, 从而 $\xi = 1$。那么，最小均方误差控制由表 13.1 中 (a) 所示的 PI 方案得到，生成的输出方差 σ_ϵ^2 为 $1.00, \sigma_x^2 = 5$。使用表 13.1 中 (b) 的最优有约束控制方程，可以使减少到 1/20(为 0.25), 而代价只是 σ_ϵ^2 提高了 20%, 达到 1.20. 但是，若用表 13.1 中简单得多的最优有约束 PI 控制器 (c), 并未付出什么代价，在两位小数的精度下就得到了同样的结果。注意，

表 13.1 对模型 (13.2.13) 和 (13.2.10) 比较不同控制方案的举例说明结果， $g=0.4, \delta=0.5, \lambda=0.4, \sigma_a^2=1$

		σ_ϵ^2	σ_x^2
(a)MMSE 控制	$-x_t=\{1+\nabla\}\epsilon_t$	1	5
(b) 最优有约束控制	$-x_t=-0.82x_{t-1}-0.21x_{t-2}$ $-0.39\epsilon_t+0.19\epsilon_{t-1}$	1.20	0.25
(c) 最优有约束 PI 控制	$-x_t=0.52\{1-0.25\nabla\}\epsilon$	1.20	0.25

如果我们用 MMSE 的 PI 方案 (a) 的人工调整图，必须从当前时刻 t 向前外推整整一个时间长度。但是对于有约束 PI 控制 (c)，我们只须从当前时刻 t 向回内插 1/4 个时间长度。这就解释了后一方案更大的稳定性。关于这个问题更完全的讨论可见 Box 和 Lucéno[53]．

13.4 对于具有固定调整和监视代价的最小代价控制

从代价角度来看，我们可以把到现在为止的讨论总结如下。如果我们假设所需要考虑的唯一代价就是对于目标的偏离，且该代价与目标偏离的平方成比例，则无约束最小均方控制就使该方案的全部代价最小．但是，另外还有一个与调整 x 的大小相联系的二次损失，并且 α 是偏离目标和作出调整的相对代价的某种度量．那么 $\sigma_\epsilon^2+\alpha\sigma_x^2$ 就是方案全部代价的度量．使这个量最小就可以得到有最小代价的控制方案，正如我们已看到的，适当选择的 PI 方案常常可以做得几乎一样好．在实际中，无论那种情况要用相对代价来估价 α 都并不容易．但是，适当方案的选择可以通过经验判断而作出，即对于 σ_ϵ^2 可接受的增大，σ_x^2 需要降低多少才能令人满意．对于具有固定调整和监视损失的系统可以使用同样的考虑．

13.4.1 对固定调整代价的有界调整

尤其是在“部件”工业中，往往会有这样的情形：一个调整可以立即生效，但是要付出固定的代价，例如，停下机器或更换工具．

有界调整图　Box 和 Jenkins 证明 [42] 在后一种情形，若假设了二次偏离目标的损失和 IMA 扰动，那么，在每次观测后都重复调整并不能达到最小损失反馈控制．与此相反，如果要求仅当对目标偏离的指数加权平均 $\hat{\epsilon}_t(1)$ 落在固定界限，比如 $\pm L$ 之外时才进行调整．我们称之为有界调整．这时应作出的调整即能

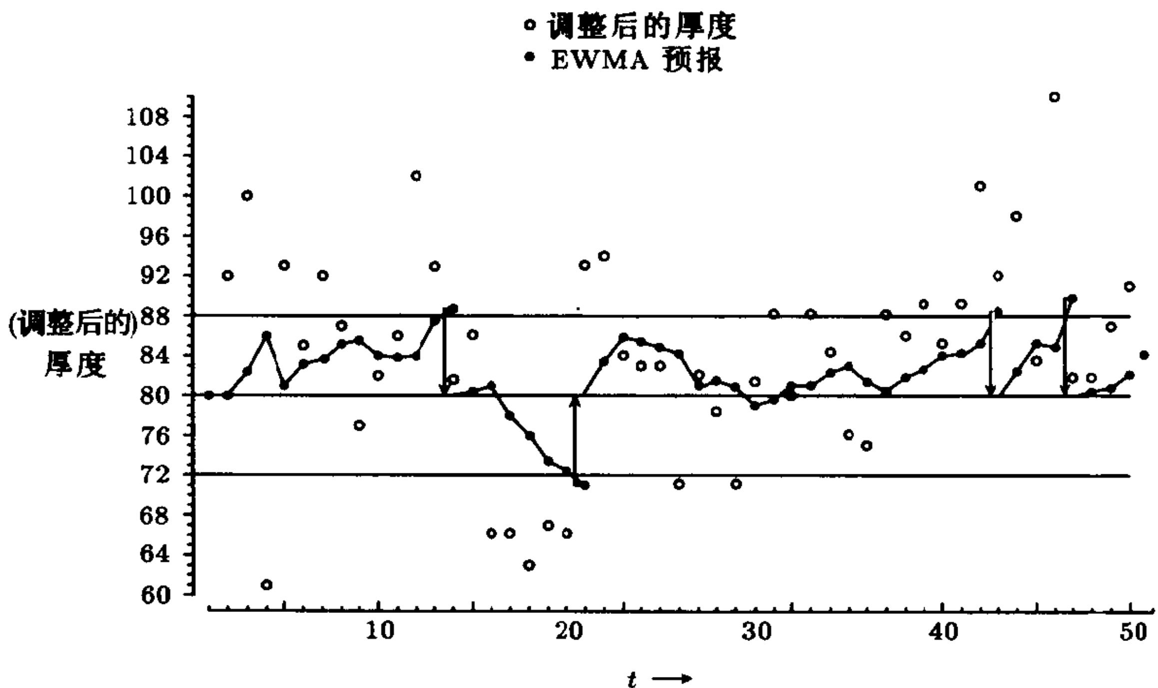


图 13.11 有限调整图：空白圈是 (调整后) 厚度偏离，实圈是这些偏离的 EWMA 预报 $\hat{\epsilon}_{t-1}(1)$

引起输出的变化 $-\hat{\epsilon}_t(1)$, 这样的调整可用如下讨论的“有界调整图”人工实现，或自动实现。

一个有界调整图　如图 13.11 所示，与 Roberts[170] 所提出的过程监视极其相似. 但是，其目的和设计是不同的。这里的目的是决定在什么时刻调整以及调整多少. 界限设计是为了使全部代价最小，它同时考虑了进行调整的代价和偏离目标的代价，它们的目的不是发现统计显著的目标偏离。当调整代价接近于零时，这两条界限越来越靠近，当调整代价为零时，它们收敛于目标值，从而得到“重复调整”的方案。

图 13.11 所示是这种图的一个例子，这是金属膜厚度的控制问题。如果改变分解速度 X 需要一个固定的代价，这就是一个合适的例子。与前面一样， $\lambda = 0.2, g = 1.2, \sigma_1 = 11$ 。在时刻 t 一个空白圆圈表示在按图形的要求定期改变分解速度 X 后所得的目标偏离 ϵ_t 。一个实圈表示适当的指数加权平均预报。它可由如

下公式方便地修正

$$\hat{\epsilon}_t(1) = \lambda\epsilon_t + \theta\hat{\epsilon}_{t-1}(1)$$

图中所示的界限为 80 ± 8, 即 $T \pm 0.72\sigma_a$. 下面我们讨论这一选择的理由. 为了理解该图是如何操作的, 假设最初分解速度是某个值 X_0 . 这个值直到 $t = 13$ 时才改变, 这时预报值为 88.7[即 $\hat{\epsilon}_t(1) = 8.7$] 落在上限之外。图形指出分解速度需要一个变化以使厚度降低 8.7. 现在对分解速度进行调整 $X_{13} - X_0 = -8.7/1.2$ 。注意, 这样的调整并不影响下一个 EWMA 的计算。例如, 时刻 $t = 14$ 时预报的厚度为 $(0.2 \times 81.3) + (0.8 \times 80.0) = 80.3$, 其中 80 是在已经作出调整以使过程回到目标之后的适当的前期预报值。

13.4.2 获得有界调整方案的间接方法

Box 和 Jenkins[42],Box,Jenkins 和 MacGregor[50] 以及 Box 和 Kramer[52] 提供了计算合适界限位置的表, 以偏离目标代价和调整代价的形式给出最小代价方案. 但是, 正如前文所述, 这些代价往往不易评估, 看起来更实际的办法是使用这些结果来提供一个最小代价方案的包, 然后按照经验从中挑选. 或者输出标准差增大, 或者调整间隔加大, 对此二者加以权衡考虑. Box[37] 对这一方法作了详细说明. 表 13.2 给出了在理论上对于各个值的 λ 和 L/σ_a, 平均调整间隔 (AAI) 和调整过程的标准差增加 (ISD), 其中有界限调整的界限为 $T \pm L$.

为了详细说明, 还是考虑金属膜厚度一例. 由 $\lambda = 0.2$, 在表 13.2 中说明了在一个有界限方案中, 对于不同选择的 L/σ_a, 标准误差增大了多少. 若 L/σ_a 置为 0.5, 标准差将增加 2.6%, 但平均起来每 10 个间隔才需做调整. 若 L/σ_a 置为 1.0, 标准差将增加 9%, 但平均调整区间 (AAI) 为 32. 图 13.22 中所绘的方案是一个折衷, 其中 L/σ_a 置为 0.72, 粗略的插值说明这将使标准差增加 5%, 而 AAI 约为 20. 为了达到这个目标, L 被置为 $8 \simeq 0.72 \times 11$. 使用图 13.2 中所绘金属厚度的 100 个观测进行蒙特卡洛方法研

究，结果表明此例中标准差实际扩大了 8.5% ,AAI 为 14. 从相对有限的样本容量来看，这种一致应视为很好的。

内插图 在应用这些理论时，可采用技术的复杂程度是任意的：从根据计算机计算采取行动的转换器，到根据简单内插图采取行动的操作员。如图 13.12 所示，它用一枚图钉和一根细线来指示适当的人工调整。在图示的情形中，时刻 $t-1$ 所作的前期预测值为 86, 而在时刻 t 所作的预测 66. 在当前时刻 t 之前，图钉在当前预报刻度上的位置是 86, 细线从钉处垂下。一旦获得实际值 66, 就拉紧细线以连接右边刻度上的点 66. 修正后的预报值

表 13.2 对于 L/σ_a 的不同选择，平均调整区间 (AAI) 和输出标准差的增加百分比 (ISD), 其中界限为 $T \pm L$

λ	L/σ_a	AAI	标准差的百分增长 ISD
0.1	0.5	32	2.4
	1.0	112	9
	1.5	243	18
	2.0	423	30
0.2	0.5	10	2.6
	1.0	32	9
	1.5	66	20
	2.0	112	32
0.3	0.5	5	2.6
	1.0	16	10
	1.5	32	20
	2.0	52	33
0.4	0.5	4	2.6
	1.0	10	10
	1.5	19	21
	2.0	32	34
0.5	0.5	3	2.5
	1.0	7	10
	1.5	13	21
	2.0	21	35

* 来源：参考文献 [37] .

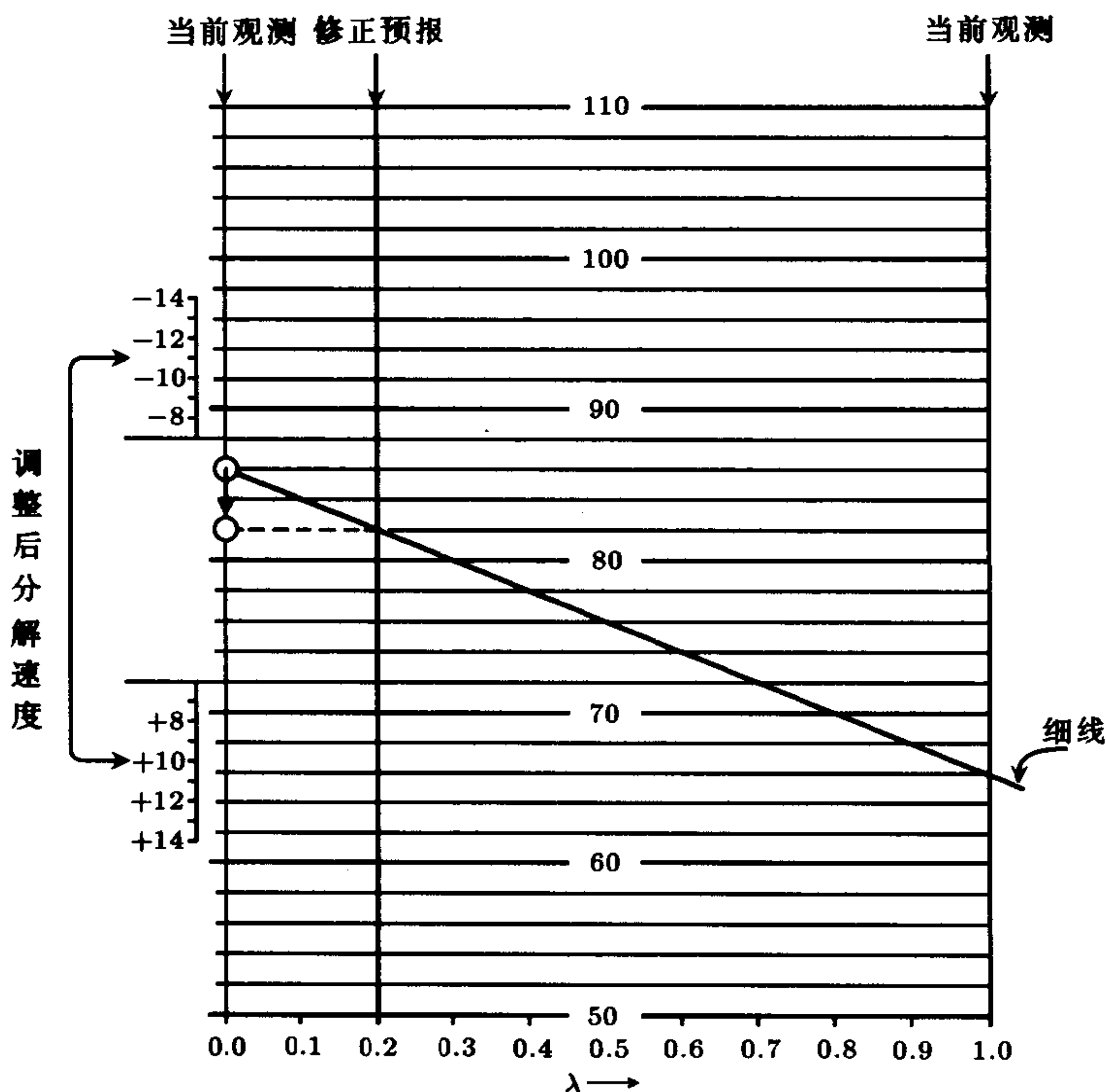


图 13.12 内插图，用于修正预报值，并指示何时分解速度应予调整及调整多少

82 可以从中间刻度上读到. 这个值是在界限内的，因此把图钉挪到这个新的当前预报值，放松细线，直到获得下一个观测，以生成新的修正预报. 一旦某个修正预报值落到某个边界之外，则应对分解速度进行适当调整，以抵消预报出的偏差，而图钉则被放到目标值以准备下一次插值.

13.4.3 监视代价的计入

Box 和 Kramer 给出了 [52] 将这些结果推广到监视代价也必

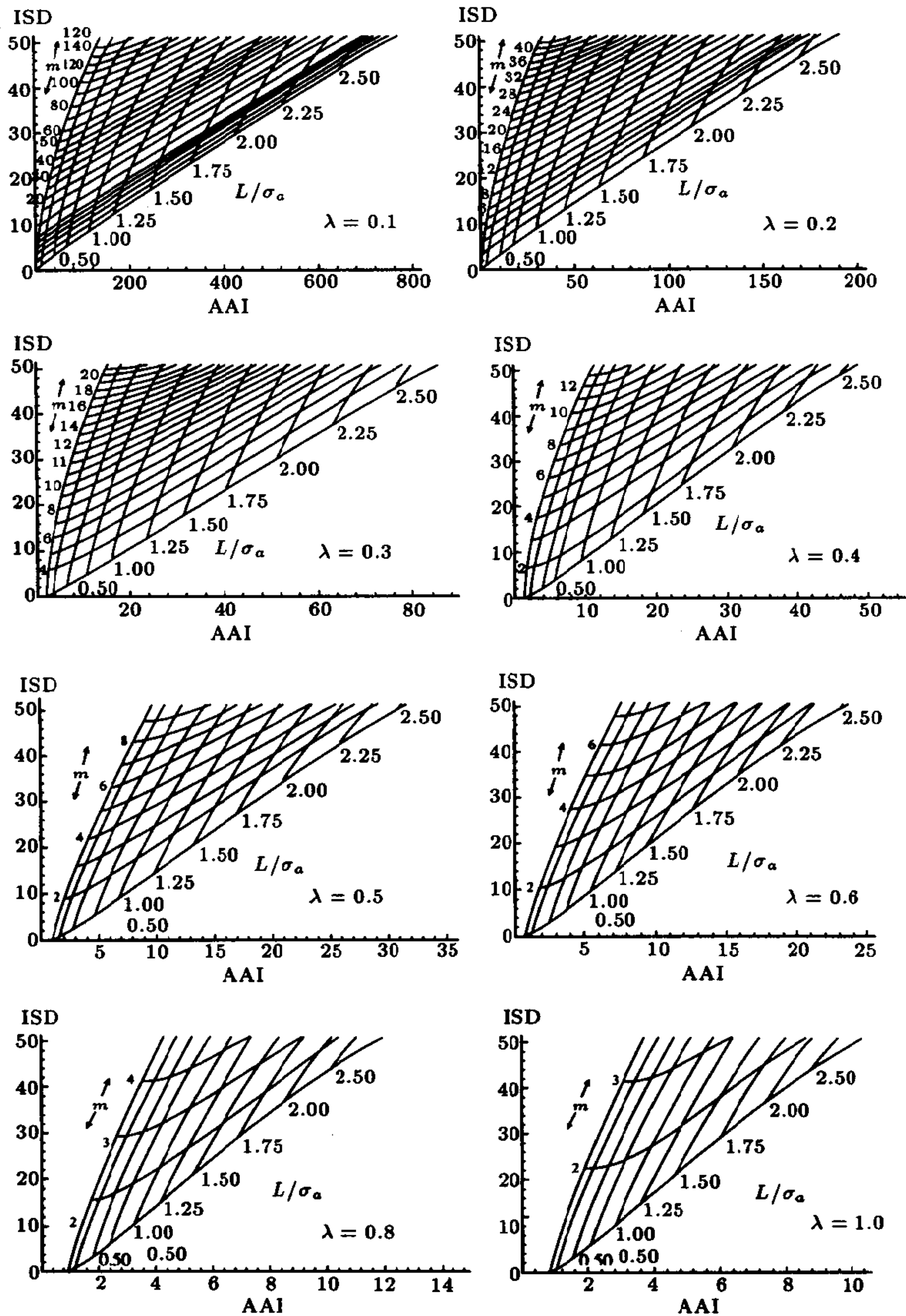


图 13.13 当 λ=0.1,0.2,0.3,0.4,0.5,0.6,0.8,1.0 时，由不同选择的 L/σ_a 和 m 得到的 AAI 和 ISD 图

须予以考虑的情形。他们考虑了进一步降低代价的可能性，即通过较少的监视 — 每 m 个间隔而不是每单位间隔监视一次。除了 σ_a 和 λ(由设备的数据来估计) 以外，给出以下 3 个代价常数是已知的： (1)(假设为二次的) 偏离目标代价， (2) 进行改变的固定代价， (3) 作出观测的固定监视代价。他们为获得最小代价方案提供了图形，给出这些信息后，得出最小代价的相应 L/σ_a 和 m 值可以由图上得到。

这三种单独的代价仍是难以确定的， Box 和 Lecéno[53] 用他们的结果允许根据经验判断选择方案。图 13.13 所示的图形给出了平均预报区间 (AAI) 和标准差增加百分数 (ISD) 的值，它关于 σ_a，其中 σ_a 对应于非平稳值的 $\lambda = 0.1(0.1) \cdot 0.6$, 0.8 和 1.0, 标准行动界限 $L/\sigma_1 = 0.0(0.25)2.5$ 及监视区间 $m = 1, 2, 3, \cdots$ 。该图包含了实际中可能需要的从小到中等的输出标准差的增加量。较大的 m 值只在 λ 较小时出现。

例如，我们先前看到用一个 $L/\sigma_a = 0.72$ 的有界调整图而不用连续方案，会使平均调整间隔增加到大约 20, 而代价只是标准差增加 5% 。这可由图 13.13 中 $\lambda = 2$ 的图形所证实，同时表明，例如，若我们监视过程的频率降低一半 $(m = 2)$, 并且仍令 $L/\sigma_a = 0.72$, 则我们可以得到相同的平均调整区间 (20), 而标准差增加 8% .

13.5 预报参数值和反馈调整方案的监视

我们先前提到了过程调整和过程监视的互补作用。这一共生现象在我们再次考虑对调整方案的监视时，需得到进一步说明．许多人提出，应该对这些方案中偏离目标的残差序列，以及类似地，对预报方案的误差加以研究，并且对残差使用 Shewart 图，或者更一般地，使用累计和或其它监视图用以预警变化。当然累计和适用于寻找均值水平的较小变化，但常常还要考虑其它种类令人担忧的偏差。Box 和 Ramirez[57](又见 Bagshaw 和 Johnson[19])

在累计 Fisher 得分统计量的基础上提出了顺序方向监视的一般理论。

假设一个模型可以写成如下形式

$$e_t = e_t(\theta) \tag{13.5.1}$$

若模型中使用了正确的参数值 $\theta = \theta_0$, 则 $\{e_t\} = \{a_t\}$ 是一个 NIID 随机变量序列. 累计得分统计量适用于检验对 θ_0 的偏离, 它可写作

$$Q = \sum e_t r_t \tag{13.5.2}$$

其中 $r_t \equiv -(de_t/d\theta)|_{\theta=\theta_0}$ 称为检测信号。

例如, 我们希望对模型 $y_t = \theta + e_t$ 检测其均值对 θ_0 的偏离。我们可以写作

$$e_t = e_t(\theta) \equiv y_t - \theta \ , \qquad a_t = y_t - \theta_0 \tag{13.5.3}$$

那么, 在本例中, 检测信号 $r_t = 1$ 和 $Q = \sum e_t$, 此即著名的累计和统计量.

一般地, 对于接近于 θ_0 的某个 θ 值, 由于 e_t 可由 $e_t = a_t - (\theta - \theta_0)r_t$ 来近似, 所以 (13.5.2) 中的累计积将包含一项

$$-(\theta - \theta_0) \sum r_t^2 \tag{13.5.4}$$

当 θ 不同于 θ_0 时, 它的大小随着样本容量 n 而系统地增加. 为了举例说明, 考虑 13.2.1 节在金属膜厚度的反馈控制方案中, 从 $t = 1$ 到 $t = 100$ 期间 λ 的值 (估计为 0.2) 发生变化的可能性. 对于这个例子

$$e_t = e_t(\theta) = \frac{1 - B}{1 - \theta B} N_t \tag{13.5.5}$$

这样

$$r_t = -\frac{1 - B}{(1 - \theta B)^2} N_{t-1} = -\frac{e_{t-1}}{1 - \theta B} = -\frac{\hat{e}_{t-1}(1)}{\lambda} \tag{13.5.6}$$

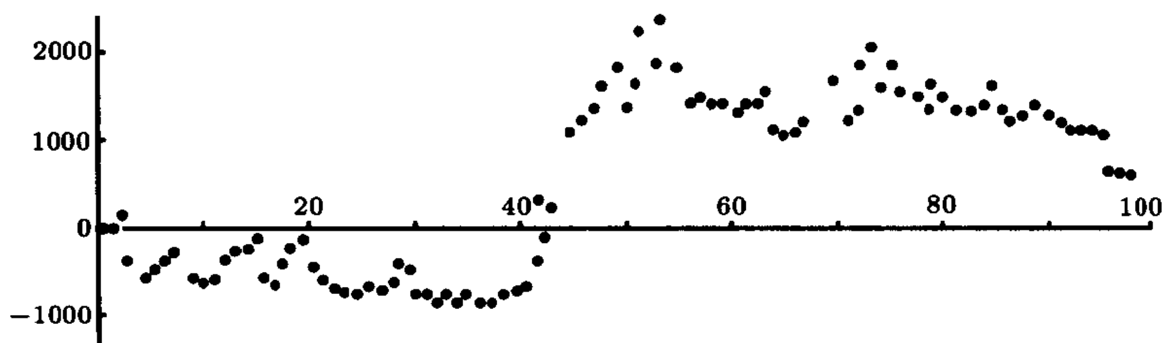


图 13.14 与图 13.3 的调整图联合使用对检测参数 θ 变化的监视

其中 $\hat{e}_{t-1}(1)=\lambda(1-\theta B)^{-1}e_{t-1}$ 是过去 e_t 的一个 EWMA 。因此检验偏离的累计得分 (Cuscore) 统计量为

$$Q=-\frac{1}{\lambda}\sum e_t\hat{e}_{t-1}(1) \tag{13.5.7}$$

其中的检测信号 $\hat{e}_{t-1}(1)$ 在本例中是残差过去值的 EWMA . 这些残差是图 13.13 中反馈调整图所绘的目标偏离. 这个准则与常识是一致的, 即如果模型为真, 则 $e_t=a_t$, 且 e_t 是不能由过去值预报的。图 13.14 所示的 Cuscore 图表明, 在大约 $t=40$ 时参数可能发生了变化。但是我们从图 13.2 的原始数据可以看到, 这非常接近于在这点上原序列的水平看来已发生变化的情形。所以这一发现需要进一步的数据和分析来确认。重要的一点在于, 这个例子说明了两类控制 (调整和监视) 和相应的两类统计推断 (估计和检验) 的互补关系. 一个进一步的发展是将滤波后的 Cuscore 统计量"馈回"控制方程, 但我们在这里对此不作讨论。

A13.1 调整方差有约束的反馈控制方案

现在考虑这样的反馈控制情形: 噪声模型和系统动态特性由 (13.2.10) 和 (13.2.13) 给出, 从而 $\epsilon_t=\mathcal{Y}_t+N_t$, 有

$$(1-B)N_t=(1-\theta B)a_t \quad 和 \quad (1-\delta B)\mathcal{Y}_t=(1-\delta)gX_{t-1+}$$

但是对输入方差 var$[x_t]$ 作一些限制是必要的，这里， $x_t=(1-B)X_t$。无约束最优方案具有这样的特征：输出误差 $\epsilon_t,\epsilon_{t-1},\epsilon_{t-2},\cdots$ 是无关随机变量 $a_t,a_{t-1},a_{t-2},\cdots$, 输出方差 σ_ϵ^2 有最小可能值 σ_a^2。对于有约束方案, 方差 σ_ϵ^2 必定高于 σ_a^2, 且输出的误差 $\epsilon_t,\epsilon_{t-1},\epsilon_{t-2},\cdots$ 会相关。

我们用如下方式提出问题：给定 σ_ϵ^2 允许增加到某个值 $\sigma_\epsilon^2=(1+c)\sigma_a^2$, 其中 c 是一个正常数。寻找得到 $\sigma_x^2=\text{var}[x_t]$ 最小值的控制方案。等价地，这个问题就是寻找表达式 $\sigma_\epsilon^2+\alpha\sigma_x^2$ 的 (无约束) 最小值， α 是某个具体的乘数，它分配 ϵ_t 和 x_t 变动的相对代价。

A13.1.1 最优调整的推导

用 a_t 的形式来表示最优调整

$$x_t=-\frac{1}{g}L(B)a_t \tag{A13.1.1}$$

其中

$$L(B)=l_0+l_1B+l_2B^2+\cdots$$

那么我们看到输出误差由下式给出

$$\begin{aligned}\epsilon_t&=\frac{(1-\delta)g}{1-\delta B}X_{t-1+}+N_t\\&=-\frac{1-\delta}{1-\delta B}(1-B)^{-1}L(B)a_t+(1-B)^{-1}(1-\theta B)a_t\\&=a_t+\left[\lambda-\frac{L(B)(1-\delta)}{1-\delta B}\right]Sa_{t-1}\end{aligned} \tag{A13.1.2}$$

其中 $S=(1-B)^{-1}$ 。表达式中 a_t 的系数是 1, 这样我们可以写

$$\epsilon_t=[1+B\mu(B)]a_t \tag{A13.1.3}$$

其中

$$\mu(B)=\mu_1+\mu_2B+\mu_3B^3+\cdots$$

而且，在实际中，控制需要用观测到的输出误差 ϵ_t 而不是 a_t 的形式来实施，因此实际用到的控制方程形式为

$$x_t = -\frac{1}{g}\frac{L(B)}{1+B\mu(B)}\epsilon_t \tag{A13.1.4}$$

使 (A13.1.2) 和 (A13.1.3) 相等，我们得到

$$(1-\delta)L(B) = [\lambda - (1-B)\mu(B)](1-\delta B) \tag{A13.1.5}$$

由于 δ, g 和 σ_a^2 是常数，所以我们可以方便地继续下去，寻找下式的无约束最小值：

$$C = \frac{(1-\delta)^2 g^2 V[x_t] + \nu V[\epsilon_t]}{\sigma_a^2} \tag{A13.1.6}$$

这里，例如有

$$V[x_t] = \text{var}[x_t]$$

及 $\nu = (1-\delta)^2 g^2/\alpha$ 。现在，由 (A13.1.3),$V[\epsilon_t]/\sigma_a^2 = 1 + \sum_{j=1}^{\infty}\mu_j^2$, 而由 (A13.1.1),$(1-\delta)gx_t = -(1-\delta)L(B)a_t = -\tau(B)a_t$, 故

$$\frac{(1-\delta)^2 g^2 V[x_t]}{\sigma_a^2} = \sum_{j=0}^{\infty}\tau_j^2$$

其中由 (A13.1.5)

$$\tau(B) = \sum_{j=0}^{\infty}\tau_j B^j = (1-\delta)L(B) = [\lambda - (1-B)\mu(B)](1-\delta B)$$

系数 $\{\tau_i\}$ 通过如下差分方程与 μ_i 函数关联：

$$\mu_i - (1+\delta)\mu_{i-1} + \delta\mu_{i-2} = -\tau_{i-1} \qquad (\text{对于} i > 2) \tag{A13.1.7}$$

这里，$\tau_0 = -(\mu_1 - \lambda), \tau_1 = -[\mu_2 - (1+\delta)\mu_1 + \lambda\delta]$。因此我们需要下式关于 μ_i 的无约束最小值

$$C = \sum_{j=0}^{\infty} \tau_j^2 + \nu \left(1 + \sum_{j=1}^{\infty} \mu_j^2\right) \tag{A13.1.8}$$

对 C 关于每个 $\mu_j(j = 1, 2, \cdots)$ 求导，令这些导数为零并求解所得的方程组，就可以获得其最小值。现在，一个给定的 μ_i 只通过 (A13.1.7) 影响 τ_{i+1}, τ_i 和 τ_{i-1}, 而且我们看到

$$\frac{\partial \tau_j}{\partial \mu_i} = \begin{cases} -1 & , \quad j = i-1 \\ 1+\delta & , \quad j = i \\ -\delta & , \quad j = i+1 \\ 0 & , \quad \text{其它} \end{cases} \tag{A13.1.9}$$

因此，由 (A13.1.8) 和 (A13.1.9) 我们得到

$$\begin{aligned} \frac{\partial}{\partial \mu_i} C &= 2\left(\tau_{i+1}\frac{\partial \tau_{i+1}}{\partial \mu_i} + \tau_i \frac{\partial \tau_i}{\partial \mu_i} + \tau_{i-1}\frac{\partial \tau_{i-1}}{\partial \mu_i} + \nu\mu_i\right) \\ &= 2[-\delta\tau_{i+1} + (1+\delta)\tau_i - \tau_{i-1} + \nu\mu_i] \qquad (\text{对于} i = 1, 2, \cdots) \end{aligned} \tag{A13.1.10}$$

然后，将 (A13.1.7) 中用 μ_i 形式表示出来的 τ_j 的表达式代入 (A13.1.10), 并令每个式子为零，我们就得到如下方程

$$i=1:\ \lambda(1+\delta+\delta^2) + 2(1+\delta+\delta^2)\mu_1 - (1+\delta)^2\mu_2 + \delta\mu_3 + \nu\mu_1 = 0 \tag{A13.1.11}$$

$$i=2:\ \lambda\delta - (1+\delta)^2\mu_1 + 2(1+\delta+\delta^2)\mu_2 - (1+\delta)^2\mu_3 + \delta\mu_4 + \nu\mu_2 = 0 \tag{A13.1.12}$$

$$i>2:\ [\delta B^2 - (1+\delta)^2 B + 2(1+\delta+\delta^2) - (1+\delta)^2 F + \delta F^2 + \nu]\mu_i = 0 \tag{A13.1.13}$$

δ 可以忽略的情形　首先考虑一种比较简单的情形，这里 δ 小得可以忽略，能够设其为零。那么上式可写为

$$i=1:\ -(\lambda - \mu_1) + (\mu_1 - \mu_2) + \nu\mu_1 = 0 \tag{A13.1.14}$$

$$i>1:\qquad [B - (2+\nu) + F]\mu_i = 0 \tag{A13.1.15}$$

这些差分方程具有如下形式的解

$$\mu_i = A_1\kappa_1^i + A_2\kappa_2^i$$

其中 κ_1 和 κ_2 是特征方程

$$B^2 - (2+\nu)B + 1 = 0 \tag{A13.1.16}$$

也即

$$B + B^{-1} = 2 + \nu$$

的根。显然，若 κ 是一个根，则 κ^{-1} 也是根。于是解的形式为 $\mu_i = A_1\kappa^i + A_2\kappa^{-i}$ 。现在，若 κ 的模小于或等于 1, 则 κ^{-1} 的模大于或等于 1, 由于 $\epsilon_t = [1 + B\mu(B)]a_t$ 的方差必须是有限的，所以对 $|\kappa| < 1$ 必有 A_2 为零。将解 $\mu_i = A_1\kappa^i$ 代入 (A13.1.14), 我们发现 $A_1 = \lambda$ 。

因此，最后有 $\mu_i = \lambda\kappa^i$, 而且由于 μ_i 和 λ 必须是实数，所以根 κ 也必须是实数。因而

$$\mu(B) = \frac{\lambda\kappa}{1-\kappa B} \qquad (0 < \kappa < 1) \tag{A13.1.17}$$

$$1 + B\mu(B) = 1 + \frac{\lambda\kappa B}{1-\kappa B} = \frac{1-\theta\kappa B}{1-\kappa B} \tag{A13.1.18}$$

这里 $\theta = 1 - \lambda$ 。于是

$$\epsilon_t = \frac{1-\theta\kappa B}{1-\kappa B}a_t$$

从而

$$\frac{V[\epsilon_t]}{\sigma_a^2} = 1 + \frac{\lambda^2\kappa^2}{1-\kappa^2} \tag{A13.1.19}$$

利用 $\delta = 0$ 的 (A13.1.5) 又有

$$L(B) = \lambda - \frac{(1-B)\lambda\kappa}{1-\kappa B} = \frac{\lambda(1-\kappa)}{1-\kappa B} \tag{A13.1.20}$$

于是

$$x_t = -\frac{\lambda}{g}\frac{(1-\kappa)}{1-\kappa B}a_t$$

和

$$\frac{V[x_t]}{\sigma_a^2} = \frac{\lambda^2}{g^2}\frac{(1-\kappa)^2}{1-\kappa^2} = \frac{\lambda^2}{g^2}\frac{1-\kappa}{1+\kappa} \tag{A13.1.21}$$

利用 (A13.1.4) 和 (A13.1.18),(A13.1.20), 我们现在找到了最优控制行动，用观测到的输出误差表示为

$$x_t = -\frac{1}{g}\frac{\lambda(1-\kappa)}{1-\theta\kappa B}\epsilon_t$$

即

$$x_t = (1-\lambda)\kappa x_{t-1} - \frac{1}{g}\lambda(1-\kappa)\epsilon_t \tag{A13.1.22}$$

注意，有约束控制方程与无约束控制方程有两处区别:

1. 引入了一个新因子 $(1-\lambda)\kappa x_{t-1}$, 从而使得当前行动部分地取决于前期行动。

2. 确定积分控制量的常数减小了，为原来的 $1-\kappa$ 倍。

我们假设可以允许输出方差增加到某个值 $\sigma_a^2(1+c)$。由 (A13.1.19) 有

$$c = \frac{\lambda^2\kappa^2}{1-\kappa}$$

即

$$\kappa = \sqrt{\frac{c}{\lambda^2+c}}$$

这里取正的平方根. 记 $Q=\kappa^2/(1-\kappa^2)$ 及 $\kappa^2 = Q/(1+Q)$，输出方差成为 $\sigma_a^2(1+\lambda^2 Q)$。

总之，假设我们准备容忍输出方差增加到 $\sigma_a^2(1+\lambda^2 Q)$, 那么

1. 我们计算 $\kappa = \sqrt{Q/(1+Q)}$。

2. 最优控制可以通过采取如下行动而获得

$$x_t = (1-\lambda)\kappa x_{t-1} - \frac{1}{g}\lambda(1-\kappa)\epsilon_t$$

3. 输入的方差将降至

$$V[x_t] = \frac{\lambda^2}{g^2}\frac{1-\kappa}{1+\kappa}\sigma_a^2$$

也即它降低后的值为无约束控制方案的 $W\%$, 这里

$$W = 100\,\frac{1-\kappa}{1+\kappa}$$

表 13.1 列出了 Q 取 0.1 到 1.0 之间值时的 κ 和 W 。例如, 假设 $\lambda = 0.4$, 这时最优无约束方案使用的控制行动为

$$x_t = -\frac{0.4}{g}\epsilon_t$$

其中 $\epsilon_t = a_t$,x_t 的方差为 $V[x_t] = (\sigma_a^2/g^2)0.16$ 。假设需要把它降低为原来的 1/4, 达到 $(\sigma_a^2/g^2)0.04$ 。这样我们要求 W 为 25% 。表 13.1 指出, 通过 $Q = 0.60$ 和 $\kappa = 0.612$ 可以把 λ 方差降至无约束值的 24% 。若我们使用这个方案, 输出方差将为

$$\sigma_\epsilon^2 = \sigma_a^2(1+0.16\times 0.60) = 1.10\sigma_a^2$$

表 A13.1 一个有约束控制方案的参数值

$c/\lambda^2=Q$	κ	W	$c/\lambda^2=Q$	κ	W
0.10	0.302	53.7	0.60	0.612	24.0
0.20	0.408	42.0	0.70	0.641	21.9
0.30	0.480	35.1	0.80	0.667	20.0
0.40	0.535	30.3	0.90	0.688	18.5
0.50	0.577	26.8	1.00	0.707	17.2

因此，使用控制行动

$$x_t = 0.37x_{t-1} - \frac{1}{g}0.16\epsilon_t$$

而不用

$$x_t = -\frac{0.4}{g}\epsilon_t$$

输入方差降为原来值的 1/4, 而输出的方差仅增加 10% 。

δ 不可忽略的情形　现在考虑更一般的情形，δ 是不可忽略的，系统的动态特性必须对它予以考虑。差分方程 (A13.1.13) 形式为

$$(\alpha B^{-2} + \beta B^{-1} + \gamma + \beta B + \alpha B^2)\mu_i = 0$$

若 κ 是特征方程的一个根. 则 κ^{-1} 也是。假设根为 $\kappa_1, \kappa_2, \kappa_1^{-1}, \kappa_2^{-1}$, 并且 κ_1 和 κ_2 是一对模小于 1 的根，那么，方程的解

$$\mu_i = A_1\kappa_1^i + A_2\kappa_2^i + A_3\kappa_1^{-i} + A_4\kappa_2^{-i}$$

A_3 和 A_4 必须为零，因为 ϵ_t 需要有有限的差分。因此，解的形式为

$$\mu_i = A_1\kappa_1^i + A_2\kappa_2^i \quad , \quad |\kappa_1| < 1, \ |\kappa_1| < 1$$

这些 A 满足 (A13.1.11) 和 (13.1.12) 所定义的初始条件，可以通过代入而得到

$$A_1 = \frac{\lambda\kappa_1(1-\kappa_2)}{\kappa_1 - \kappa_2}, \qquad A_2 = -\frac{\lambda\kappa_2(1-\kappa_1)}{\kappa_1 - \kappa_2}$$

如果我们记 $k_0 = \kappa_1 + \kappa_2 - \kappa_1\kappa_2, k_1 = \kappa_1\kappa_2$, 于是有

$$\mu(B) = \lambda\left[\frac{k_0 - k_1 B}{1 - (k_0 + k_1)B + k_1 B^2}\right] \qquad \text{(A13.1.23)}$$

和

$$1+B\mu(B)=\frac{1-k_1B-(1-\lambda)(k_0B-k_1B^2)}{1-(k_0+k_1)B+k_1B^2} \qquad \text{(A13.1.24)}$$

现在把 (A13.1.23) 代入 (A13.1.5),

$$L(B)=\frac{\lambda(1-\delta B)(1-k_0)}{(1-\delta)[1-(k_0+k_1)B+k_1B^2]} \qquad \text{(A13.1.25)}$$

及

$$\frac{L(B)}{1+B\mu(B)}=\frac{\lambda(1-\delta B)(1-k_0)}{(1-\delta)[1-k_1B-(1-\lambda)(k_0B-k_1B^2)]}$$

因此，由 (A13.1.4) 我们找到了用误差 ϵ_t 的形式来表示的最优控制行动

$$x_t=-\frac{\lambda}{g}\frac{(1-\delta B)(1-k_0)}{(1-\delta)[1-k_1B-(1-\lambda)(k_0B-k_1B^2)]}\epsilon_t \qquad \text{(A13.1.26)}$$

或

$$x_t=[k_1+(1-\lambda)k_0]x_{t-1}-(1-\lambda)k_1x_{t-2}-\frac{\lambda(1-k_0)(1-\delta B)}{g(1-\delta)}\epsilon_t \qquad \text{(A13.1.27)}$$

这样，修正后的控制方案使得 x_t 依赖于 x_{t-1} 和 x_{t-2}(若 $\lambda=1$ 则只依赖于 x_{t-1}), 并且将标准积分比例行动减小到原来的 $1-k_0$ 倍.

输出和输入的方差 输入和输出的实际方差很容易得出，因为

$$\epsilon_t=a_t+\lambda\left[\frac{k_0-k_1B}{1-(k_0+k_1)B+k_1B^2}\right]a_{t-1}$$

等式右端的第二项定义了一个阶数为 (2,0,1) 的混合自回归滑动平均过程， ϵ_t 的方差可得出为

$$\frac{V[\epsilon_t]}{\sigma_a^2}=1+\lambda^2\left\{\frac{(k_0+k_1)^2(1-k_1)-2k_1(k_0-k_1^2)}{(1-k_1)[(1+k_1)^2-(k_0+k_1)^2]}\right\}=1+\lambda^2 Q \tag{A13.1.28}$$

还有

$$\frac{V[x_t]}{\sigma_a^2}=\frac{\lambda^2}{g^2(1-\delta)^2}\frac{(1-k_0)[(1+\delta^2)(1+k_1)-2\delta(k_0+k_1)]}{(1+k_0+2k_1)(1-k_1)} \tag{A13.1.29}$$

计算 k_0 和 k_1 　回到差分方程 (A13.1.13), 其特征方程可写为

$$B^4-MB^3+NB^2-MB+1=0$$

其中 $M=(1+\delta)^2/\delta, N=[(1+\delta)^2+(1+\delta^2)+\nu]/\delta$ 。它还可以写成

$$(B^2-TB+P)(B^2-P^{-1}TB+P^{-1})=0$$

这里

$$T=\kappa_1+\kappa_2 \qquad 和 \qquad P=\kappa_1\kappa_2$$

令 B 的系数相等，我们得到

$$T+P^{-1}T=M$$

即 $T=PM/(1+P)$, 而

$$P+P^{-1}+P^{-1}T^2=N$$

这样 $P+P^{-1}+PM^2/(1+P)^2=N$, 即

$$(P+2+P^{-1})(P+P^{-1})+M^2=N(P+2+P^{-1})$$

或

$$(P+P^{-1})^2+(2-N)(P+P^{-1})+M^2-2N=0$$

对于适当的 ν 值，这个二次方程有两个实根

$$u_1 = \kappa_1\kappa_2 + \kappa_1^{-1}\kappa_2^{-1} \qquad u_2 = \kappa_1\kappa_2^{-1} + \kappa_1^{-1}\kappa_2$$

u_1 是较大的根。所要求的 P 现在是如下二次方程较小的根

$$P^2 - u_1P + 1 = 0$$

T 由下式给出

$$T = [P(u_2+2)]^{1/2}$$

有约束方案的最优值表

表的构造　表 13.2 是为便于最优控制方案选择而提供的。表中的值是对于传递函数模型中每个选定的参数 δ 值用如下方式得到的：

1. 计算$M = \dfrac{(1+\delta)^2}{\delta}$和 $N = \dfrac{(1+\delta)^2 + (1+\delta^2) + \nu}{\delta}$,
   以便为对一系列ν值的挑选提供Q的适当范围。
2. 计算$u_1 = \dfrac{1}{2}(N-2) + \left[\left(\dfrac{N-2}{2}\right) + 2N - M^2\right]^{1/2}$
   和$u_2 = \dfrac{1}{2}(N-2) - \left[\left(\dfrac{N-2}{2}\right) + 2N - M^2\right]^{1/2}$。
3. 计算$k_1 = P = \dfrac{1}{2}u_1 - \left[\left(\dfrac{1}{2}u_1\right)^2 - 1\right]^{1/2}$。
   和$k_0 = T - P = [k_1(u_2+2)]^{1/2} - k_1$.
4. 计算$Q = \dfrac{(k_0+k_1)^2(1-k_1) - 2k_1(k_0-k_1^2)}{(1-k_1)[(1+k_1)^2 - (k_0+k_1)^2]}$。
5. 计算$W = \dfrac{(1-k_0)[(1+\delta^2)(1+k_1) - 2\delta(k_0+k_1)]}{(1+k_0+2k_1)(1-k_1)(1+\delta^2)}$。
6. 在Q的适当值处，内插W, k_0和k_1.

表 A13.2 用于方便计算最优有约束控制方案的表

δ		100Q				
		20	40	60	80	100
0.9	100W	21.7	11.3	6.7	4.5	3.1
	k_0	0.44	0.585	0.68	0.74	0.78
	k_1	0.18	0.27	0.34	0.39	0.44
0.8	100W	22.0	11.7	7.2	4.8	3.4
	k_0	0.44	0.585	0.68	0.74	0.78
	k_1	0.18	0.27	0.33	0.38	0.43
0.7	100W	22.7	12.4	8.0	5.6	4.1
	k_0	0.44	0.585	0.68	0.74	0.78
	k_1	0.17	0.25	0.32	0.36	0.40
0.6	100W	24.1	13.6	9.0	6.6	5.0
	k_0	0.44	0.58	0.67	0.73	0.78
	k_1	0.16	0.24	0.29	0.33	0.365
0.5	100W	26.5	15.5	10.5	7.9	6.2
	k_0	0.43	0.58	0.67	0.72	0.77
	k_1	0.15	0.21	0.26	0.29	0.32
0.4	100W	28.5	17.7	12.7	9.8	7.9
	k_0	0.43	0.57	0.66	0.72	0.76
	k_1	0.13	0.18	0.22	0.245	0.265
0.3	100W	31.5	20.5	15.2	12.0	9.9
	k_0	0.43	0.57	0.65	0.71	0.75
	k_1	0.105	0.145	0.17	0.19	0.20
0.2	100W	34.8	23.6	18.0	14.5	12.2
	k_0	0.42	0.56	0.64	0.69	0.73
	k_1	0.07	0.10	0.12	0.13	0.14
0.1	100W	38.2	26.7	21.0	17.3	14.6
	k_0	0.42	0.55	0.63	0.68	0.72
	k_1	0.04	0.05	0.06	0.065	0.07

表的使用　表 A13.2 可以使用如下：δ 的值列在纵栏第一列。由 $V[\epsilon_t]=(1+\lambda^2 Q)\sigma_a^2$ 这一事实，输出方差增加的百分数是 $100Q\lambda^2$ 。适当的 Q 值列在横栏第一行。表中的元素为 (1) $100W$，是 x_t 降低的百分数，(2)k_0 和 (3)k_1 。

例如，假设 $\lambda=0.6, \delta=0.5, g=1$ 。那么最优无约束控制方

程为

$$x_t = -1.2(1-0.5B)\epsilon_t = -1.2(1-0.5B)a_t$$

及 $\mathrm{var}[x_t] = 1.80\sigma_a^2$ 。假设输入变量的这一变动幅度在过程运行中带来了一些困难，需要将 $\mathrm{var}[x_t]$ 降低到 $0.50\sigma_a^2$, 即无约束方案值的约 28% 。查阅表 A13.2,$\delta = 0.5$ 所标的这一行指出，通过使用常数 $k_0 = 0.43, k_1 = 0.15$ 的控制方案，即使用控制方程 (A13.1.27) 给出

$$x_t = 0.32x_{t-1} - 0.06x_{t-2} - (0.57\times 1.2)(1-0.5B)\epsilon_t$$

能达到降低 26.5% 。这个解对应于 $Q = 0.20$. 因此输出方差将增加的倍数是 $1+\lambda^2 Q = 1 + 0.6^2(0.2) = 1.072$, 即约 7% 。

A 13.2 采样间隔的选择

与连续系统比较起来，如果恰当地选择采样间隔，我们这里讨论的离散控制系统可以是非常有效的。粗略地说，我们需要的间隔应该使得在采样间隔内不会发生过大的变化。通常，那些贯穿全部或部分系统的扰动特征反映了系统的惯性和动态特性，所以采样间隔常常是试探性地选取，或直接与时间常数或系统常数成比例。在涉及液体反应和混合的化学过程中，以小时计的间隔和操作员监视以及人工调整可能就足够了。相比之下，在涉及气体反应时，恰当选择的采样间隔可能是以秒计的，而且需要自动监视和调整。

在一些场合，可能需要用试验来得到令人满意的采样间隔，而在其它情形中，相当简单的计算就能指出采样间隔的选择将如何影响可能的控制程度。

A13.2.1 降低采样频率的效果示例

为了详细说明这类有意义的计算，仍然假设我们有一个简单的系统，并使用某一特定的采样间隔，噪声由一个 (0,1,1) 过程

$\nabla N_t = (1-\theta B)a_t$ 表示，传递函数模型由一阶系统 $(1-\delta B)\mathcal{Y}_t = g(1-\delta)x_{t-1}$ 表示。在此例中，如果我们使用 MMSE 调整

$$x_t = \frac{1-\theta}{g(1-\delta)}(1-\delta B)\epsilon_t \qquad \text{(A13.2.1)}$$

那么目标偏离是 $\epsilon_t = a_t$, 假设有方差 $\sigma_a^2 = \sigma_1^2$,

实际中，常常提出这样的问题：如果采样频率再低一些，我们将会损失多少？为了回答这个问题，我们考虑对所涉及随机过程的采样效果。

A13.2.2 一个 IMA(0,1,1) 过程的采样

假设对于由在某“单位”间隔作出的观测，我们有一个噪声模型

$$\nabla N_t = (1-\theta_1 B)a_t$$

$\mathrm{var}[a_t] = \sigma_a^2 = \sigma_1^2$, 其中的下标 1 在这里用来表示采样间隔的选择。那么，对于差分 ∇N_t, 其自相关 γ_k 由下式给出

$$\begin{aligned} \gamma_0 &= (1+\theta_1^2)\sigma_1^2 \\ \gamma_1 &= -\theta_1\sigma_1^2 \\ \gamma_j &= 0 \qquad (j > 2) \end{aligned} \qquad \text{(A13.2.2)}$$

记 $\zeta = (\gamma_0 + 2\gamma_1)/\gamma_1$, 我们得到

$$\zeta = -\frac{(1-\theta_1)^2}{\theta_1}$$

这样，给定 γ_0 和 γ_1, 通过解二次方程

$$(1-\theta_1)^2 - \zeta(1-\theta_1) + \zeta = 0$$

选择 $-1 < \theta_1 < 1$ 的根，就可以得到 IMA 过程的参数 λ 。而且

$$\sigma_1^2 = -\frac{\gamma_1}{\gamma_0} \qquad \text{(A13.2.3)}$$

现在假设在 h 个单位的间隔观测过程 N_t(h 是一个正整数), 所得到的过程用 M_t 表示。那么

$$\begin{aligned}\nabla M_t &= N_t - N_{t-h} \\ &= (a_t + a_{t-1} + \cdots + a_{t-h+1}) - \theta_1(a_{t-1} + a_{t-2} + \cdots + a_{t-h}) \\ \nabla M_{t-h} &= N_{t-h} - N_{t-2h} \\ &= (a_{t-h} + a_{t-h-1} + \cdots + a_{t-2h+1}) - \theta_1(a_{t-h-1} + \cdots + a_{t-2h})\end{aligned}$$

等等。对于差分 ∇M_t, 自相关 $\gamma_k(h)$ 为

$$\begin{aligned}\gamma_0(h) &= [(1+\theta_1^2) + (h-1)(1-\theta_1)^2]\sigma_1^2 \\ \gamma_1(h) &= -\theta_1\sigma_1^2 \\ \gamma_j(h) &= 0 \qquad (j > 2)\end{aligned} \tag{A13.2.4}$$

那么过程 M_t 也是一个阶数为 (0,1,1) 的 IMA 过程,

$$\nabla M_t = (1 - \theta_h B)e_t$$

其中 e_t 是白噪声过程, 方差为 σ_h^2 。现在

$$\frac{\gamma_0(h) + 2\gamma_1(h)}{\gamma_1(h)} = -\frac{h(1-\theta_1)^2}{\theta_1}$$

从而

$$\frac{h(1-\theta_1)^2}{\theta_1} = \frac{(1-\theta_h)^2}{\theta_h} \tag{A13.2.5}$$

而且, 由于 $\gamma_1(h) = -\theta_h\sigma_h^2 = -\theta_1\sigma_1^2$, 有

$$\frac{\sigma_h^2}{\sigma_1^2} = \frac{\theta_1}{\theta_h} \tag{A13.2.6}$$

因此, 我们已经证明了以间隔 h 对一个阶数为 (0,1,1) 的 IMA 过程采样, 得出了另一个阶数为 (0,1,1) 的 IMA 过程。由 (A13.2.5) 我们可以得到采出过程的参数 θ_h 值, 由 (A13.2.6), 我们可以得到用过程参数 θ_1 和 σ_a^2 表示的该过程方差 σ_h^2 。

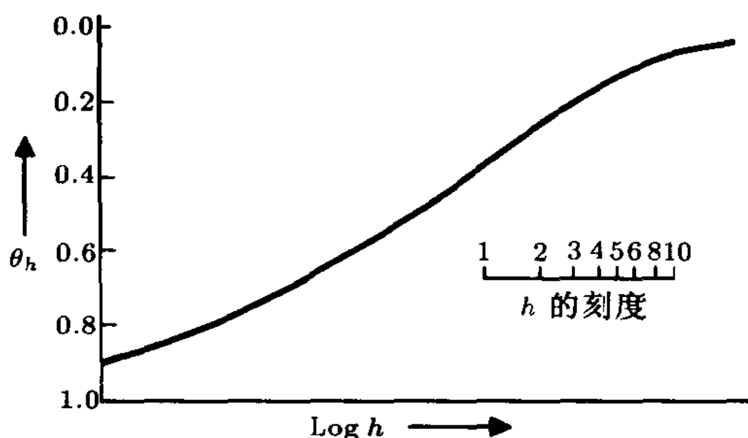


图 A13.1 对 IMA(0,1,1) 过程采样，θ_h 相对于 $\log h$ 给出

图 A13.1 中，θ_h 相对于 $\log h$ 而绘出，还补充了一个 h 的比例尺。这个图可以使人们找到对一个 (0,1,1) 过程任意扩大采样间隔的效果。例如，假设我们有一个 $\theta_1 = 0.5, \sigma_1^2 = 1$ 的过程。让我们用这个图来寻找采样间隔 (a) 翻倍，(b) 翻两倍 —相应参数 $\theta_2, \theta_4, \sigma_2^2\sigma_4^2$ 的值。在一张纸的边缘用图中的刻度标出 $h = 1, h = 2$ 和 $h = 4$ 的点，我们把这张纸水平地放置，使得 $h = 1$ 在曲线上对应于点 $\theta_1 = 0.5$ 。然后我们可以读出对应于 $h = 2$ 和 $h = 4$ 的 θ_2 和 θ_4 的坐标。我们发现

$$\theta_1 = 0.5 \quad , \quad \theta_2 = 0.38 \quad , \quad \theta_3 = 0.27$$

使用 (A13.2.6), 方差与 θ 值成反比，从而有

$$\sigma_1^2 = 1.00 \quad , \quad \sigma_2^2 = 1.32 \quad , \quad \sigma_4^2 = 2.17$$

现在假设对于单位间隔的原方案，动态常数为 δ_1(我们仍将用这个下标表示采样间隔) 。那么，由于在实际时间中，同一个固定的时间常数 $T = -h/\ln(\delta)$ 应用于所有的方案，所以我们有

$$\delta_2 = \delta_1^2 \quad , \quad \delta_4 = \delta_1^4$$

对于某一特定采样间隔 h 给出最小均方误差的方案为

$$x_t(h) = \frac{1 - \theta_h}{g(1 - \delta_1^h)}(1 - \delta_1^h B)\epsilon_t(h)$$

或

$$x_t(h) = -\frac{1-\theta_h}{g}\left(1+\frac{\delta_1^h}{1-\delta_1^h}\nabla\right)\epsilon_t(h) \qquad \text{(A13.2.7)}$$

假设，例如和上面一样， $\theta_1 = 0.5, \delta_1 = 0.8$, 所以 $\delta_2 = 0.64, \delta_4 = 0.41$ 。那么最优方案为

$$h=1: x_t(1) = -\frac{0.5}{g}(1+4\nabla)\epsilon_t(1) \qquad , \sigma_\epsilon^2 = 1.00, g^2\sigma_x^2 = 10.25$$

$$h=2: x_t(2) = -\frac{0.62}{g}(1+1.78\nabla)\epsilon_t(2) \quad , \sigma_\epsilon^2 = 1.32, g^2\sigma_x^2 = 5.50$$

$$h=4: x_t(4) = -\frac{0.73}{g}(1+0.69\nabla)\epsilon_t(4) \quad , \sigma_\epsilon^2 = 2.17, g^2\sigma_x^2 = 3.84$$

与料想一致，随着采样间隔的扩大和系统动态重要性的降低，“积分”控制的量增加，而且比例控制与积分控制的比率显著降低。我们先前注意到过大的调整方差通常是不利的。所列的 $g\sigma_x^2$ 值表示了这些方案在这一点上如何不同。当然，σ_x^2 的较小值自身并不能修正 $h=4$ 的选择. 用一个最优有约束方案，比如附录 A13.1 中所述的， $h=1$, 可以带来 σ_x^2 很大的降低而输出方差只增加很少。例如，以 $\delta = 0.8, 100Q = 20$ 带入 A13.2 表中，我们发现对于输出方差增加 5% 达到 $(1+\lambda^2\theta)\sigma_1^2 = 1.05\sigma_1^2$ 时， $h=1$ 方案的输入方差将降低到其无约束值的 22%, 从而 $g^2\sigma_x^2 = 10.25 \times 0.22 = 2.26$. 使用 (A13.1.27), 我们得到对于 $h=1$ 的有约束方案

$$x_t = 0.40x_{t-1} - 0.09x_{t-2} - 0.56\left[\frac{0.5}{g}(1+4\nabla)\right]\epsilon_t(1)$$

$$\sigma_\epsilon^2 = 1.05 \qquad , \quad g^2\sigma_x^2 = 2.26$$

在实际中，可以设置各种替换的方案及其伴随的特征，从中作出经济的选择来解决具体问题。一般地，由扩大采样间隔引起的输出方差的增加，必须由低频监视可能带来的经济上的利益来予以补偿。

第五部分
图　表

本书的这一部分是对时间序列分析很有用的辅助资料汇集，其中包括了为得到自回归滑动平均模型初估计的许多表和图，同时还有常用的正态分布，χ^2 分布，t 分布等的尾部面积表. 其后完整地列出了本书中分析的所有时间序列，在第六部分的习题和问题中所讨论的另一些时间序列也列出了.

图表汇集

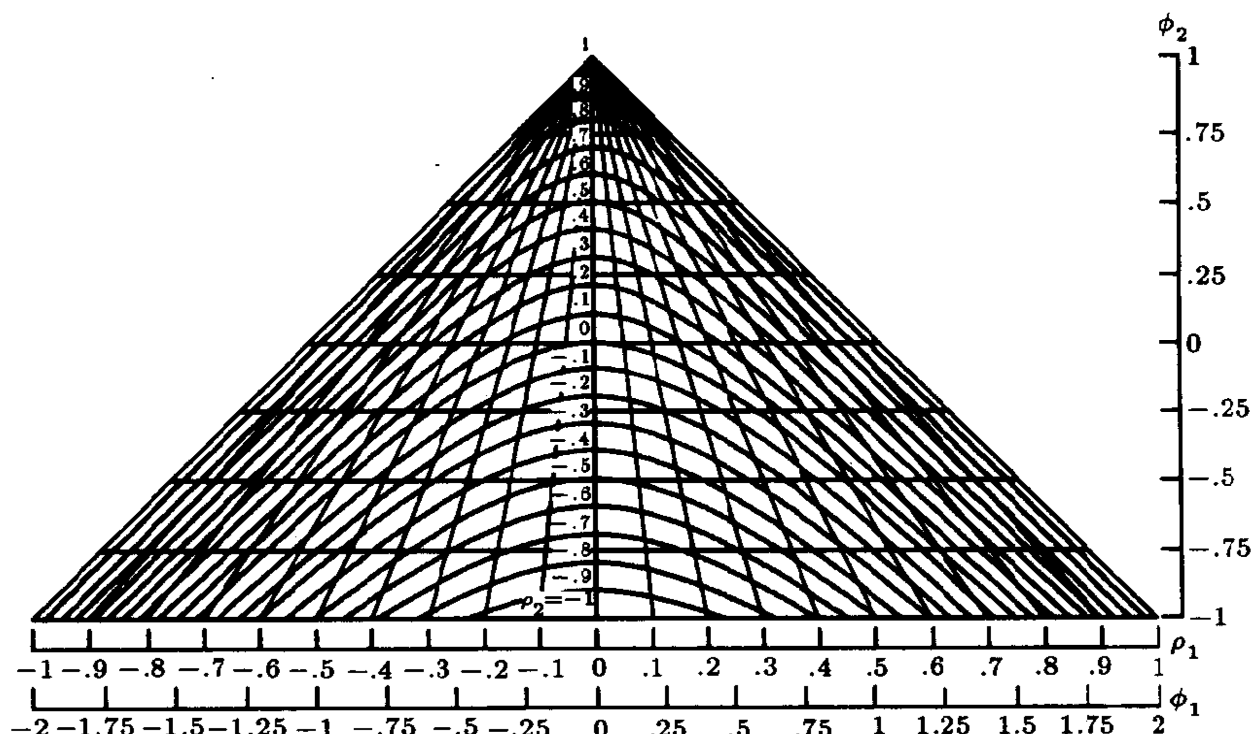
图 B,C 和 D 的采用和复制是经 [183] 作者的许可。表 E,F, 和 G 引自统计学，卷 I 的生物统计表，这些表的压缩和采用是经生物统计理事会的许可。

表 A　一阶滑动平均过程 ρ_1 关于 θ 的表

θ	ρ_1	θ	ρ_1	θ	ρ_1	θ	ρ_1
0.00	0.000	0.55	–0.422	0.00	0.000	–0.55	0.422
0.05	–0.050	0.60	–0.441	–0.05	0.050	–0.60	0.441
0.10	–0.099	0.65	–0.457	–0.10	0.099	–0.65	0.457
0.15	–0.147	0.70	–0.468	–0.15	0.147	–0.70	0.468
0.20	–0.192	0.75	–0.480	–0.20	0.192	–0.75	0.480
0.25	–0.235	0.80	–0.488	–0.25	0.235	–0.80	0.488
0.30	–0.275	0.85	–0.493	–0.30	0.275	–0.85	0.493
0.35	–0.315	0.90	–0.497	–0.35	0.315	–0.90	0.497
0.40	–0.349	0.95	–0.499	–0.40	0.349	–0.95	0.499
0.45	–0.374	1.00	–0.500	–0.45	0.374	–1.00	0.500
0.50	–0.400			–0.50	0.400		

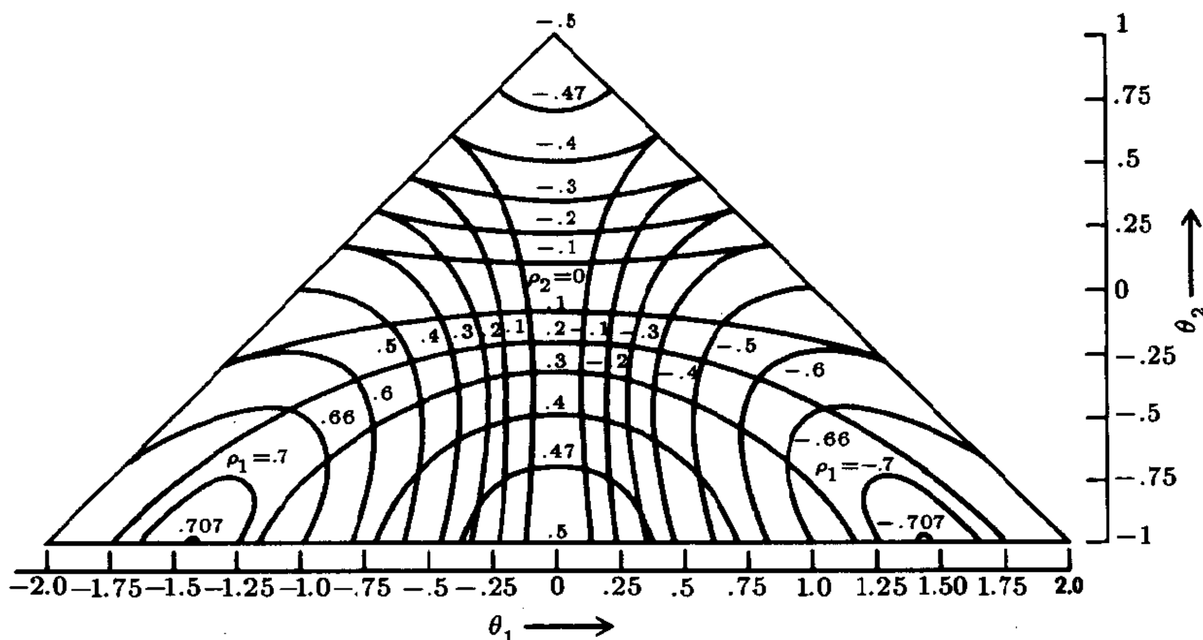
注：该表可以用来得到 (0,d,1) 过程 $w_t=(1-\theta B)a_t$ 参数的初估计，其中 $w_t=\nabla^d z_t$，且用 $r_1(w)$ 代替 ρ_1 。

图 B　二阶自回归过程 ρ_1 和 ρ_2 关于 ϕ_1 和 ϕ_2 的图



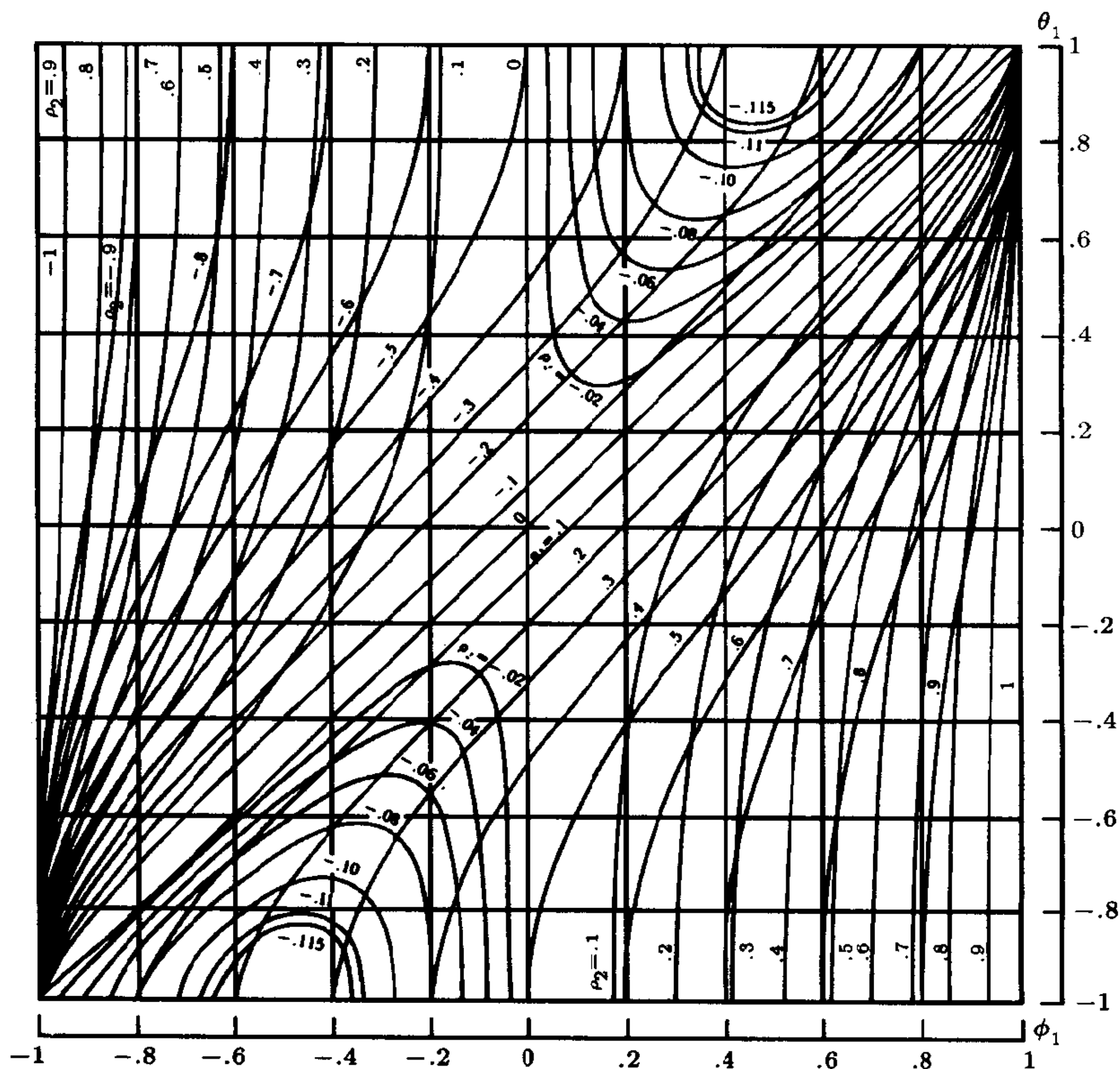
注：该图可以用来得到 $(2,d,0)$ 过程 $(1-\phi_1B-\phi_2B^2)w_t=a_t$ 参数的估计值，其中 $w_t=\nabla^d z_t$，且用 $r_1(w)$ 和 $r_2(w)$ 代替 ρ_1 和 ρ_2．

图 C　二阶滑动平均过程 ρ_1 和 ρ_2 关于 θ_1 和 θ_2 的图



注：该图可以用来得到 $(0,d,2)$ 过程 $w_t=(1-\theta_1B-\theta B^2)a_t$ 参数的估计值，其中 $w_t=\nabla^d z_t$，且用 $r_1(w)$ 和 $r_2(w)$ 代替 ρ_1 和 ρ_2．

图 D　一阶自回归－一阶滑动平均混合过程 ρ_1 和 ρ_2 关于 ϕ_1 和 θ_1 的图



注：该图可以用来得到 $(1,d,1)$ 过程 $(1-\phi B)w_t=(1-\theta_1 B)a_t$ 参数的估计值，其中 $w_t=\nabla^d z_t$，且用 $r_1(w)$ 和 $r_2(w)$ 代替 ρ_1 和 ρ_2 .

表 E　标准正态分部布的尾部面积

u_ϵ	ϵ	$p(u_\epsilon)$	u_ϵ	ϵ	$p(u_\epsilon)$
0.0	0.500	0.3989	1.6	0.055	0.1109
0.1	0.460	0.3969	1.7	0.045	0.094
0.2	0.421	0.3910	1.8	0.036	0.0790
0.3	0.382	0.3814	1.9	0.029	0.0656
0.4	0.345	0.3683	2.0	0.023	0.0540
0.5	0.309	0.3521	2.1	0.018	0.0440
0.6	0.274	0.3322	2.2	0.014	0.0355
0.7	0.242	0.3123	2.3	0.011	0.0283
0.8	0.212	0.2897	2.4	0.008	0.0224
0.9	0.184	0.2661	2.5	0.006	0.0175
1.0	0.159	0.2420	2.6	0.005	0.0136
1.1	0.136	0.2179	2.7	0.003	0.0104
1.2	0.115	0.1942	2.8	0.003	0.0079
1.3	0.097	0.1714	2.9	0.002	0.0059
1.4	0.081	0.1497	3.0	0.001	0.0044
1.5	0.067	0.1295			

注：表中列出了单位正态的偏差 u_x，它使 $\Pr\{u_> u_\epsilon\} = \epsilon$，也列出了纵坐标值 $p(u = u_\epsilon)$.

表 F χ^2 分布的尾部面积

	ϵ														
m	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001	m
1	—	—	—	—	0.016	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88	10.8	1
2	0.010	0.020	0.051	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	13.8	2
3	0.072	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	16.3	3
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	18.5	4
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	20.5	5
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5	22.5	6
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3	24.3	7
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0	26.1	8
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6	27.9	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2	29.6	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	17.3	13.7	19.7	21.9	24.7	26.8	31.3	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3	32.9	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8	34.5	13
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3	36.1	14
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8	37.7	15
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3	39.3	16
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7	40.8	17
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2	42.3	18
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6	43.8	19
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0	45.3	20
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4	46.8	21
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8	48.3	22
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2	49.7	23
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.9	45.6	51.2	24
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9	52.6	25
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	54.1	26
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	55.5	27
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	56.9	28
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	58.3	29
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	59.7	30

注：表中列出了 $\chi_\epsilon^2(m)$ 的值，使得 $\Pr\{(\chi^2(m)>\chi_\epsilon^2(m)\}=\epsilon$，其中 m 为自由度个数。

表 G　t 分布的尾部面积

	ϵ					
ν	0.25	0.10	0.05	0.025	0.01	0.005
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.96	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.06
5	0.73	1.48	2.02	2.57	3.36	4.03
6	0.72	1.44	1.94	2.45	3.14	3.71
7	0.71	1.42	1.90	2.36	3.00	3.50
8	0.71	1.40	1.86	2.31	2.90	3.36
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11
12	0.70	1.36	1.78	2.18	2.68	3.06
13	0.69	1.35	1.77	2.16	2.65	3.01
14	0.69	1.34	1.76	2.14	2.62	3.00
15	0.69	1.34	1.75	2.13	2.60	2.95
16	0.69	1.34	1.75	2.12	2.58	2.92
17	0.69	1.33	1.74	2.11	2.57	2.90
18	0.69	1.33	1.73	2.10	2.55	2.88
19	0.69	1.33	1.73	2.09	2.54	2.86
20	0.69	1.33	1.72	2.09	2.53	2.84
30	0.68	1.31	1.70	2.04	2.46	2.75
40	0.68	1.30	1.68	2.02	2.42	2.70
60	0.68	1.30	1.67	2.00	2.39	2.66
120	0.68	1.29	1.66	1.98	2.36	2.62
∞	0.67	1.28	1.64	1.96	2.33	2.58

注：表中列出了 $t_\epsilon^2(\nu)$ 的值，使得 $\Pr\{t(\nu)>t_\epsilon(\nu)\}$，其中 ν 为自由度个数.

正文和习题中使用的时间序列汇集

序列 A	化工生产过程的浓度读数 (每 2 小时)
序列 B	IBM 普通股收盘价 (1961.5.17—1962.11.2, 每日)
序列 B′	IBM 普通股收盘价 (1959.6.29—1960.6.30, 每日)
序列 C	化工生产过程的温度读数 (每分钟)
序列 D	化工生产过程的粘度读数 (每小时)
序列 E	Wölfer 太阳黑子数 (每年)
序列 F	化工生产过程的产量 (顺次)
序列 G	国际航线旅客数 (1949.01—1960.12, 每月总数, 单位: 千人)
序列 J	煤气炉数据
序列 K	两输入的动态模拟数据
序列 L	试验方案数据
序列 M	具有领先指标的销售额数据
序列 N	哈德逊湾 (Husdon Bay) 公司貂皮销售量数据 (1850–1911, 每年)
序列 P	英国的失业和 GDP 数据 (1955—1969, 每季度)
序列 Q	美国生猪价格数据 (1867—1948, 每年)
序列 R	洛杉基市臭氧每小时读数的月平均值

序列 A 化工生产过程的浓度读数 *

(每 2 小时)

1	17.0	41	17.6	81	16.8	121	16.9	161	17.1
2	16.6	42	17.5	82	16.7	122	17.1	162	17.1
3	16.3	43	16.5	83	16.4	123	16.8	163	17.1
4	16.1	44	17.8	84	16.5	124	17.0	164	17.4
5	17.1	45	17.3	85	16.4	125	17.2	165	17.2
6	16.9	46	17.3	86	16.6	126	17.3	166	16.9
7	16.8	47	17.1	87	16.5	127	17.2	167	16.9
8	17.4	48	17.4	88	16.7	128	17.3	168	17.0
9	17.1	49	16.9	89	16.4	129	17.2	169	16.7
10	17.0	50	17.3	90	16.4	130	17.2	170	16.9
11	16.7	51	17.6	91	16.2	131	17.5	171	17.3
12	17.4	52	16.9	92	16.4	132	16.9	172	17.8
13	17.2	53	16.7	93	16.3	133	16.9	173	17.8
14	17.4	54	16.8	94	16.4	134	16.9	174	17.6
15	17.4	55	16.8	95	17.0	135	17.0	175	17.5
16	17.0	56	17.2	96	16.9	136	16.5	176	17.0
17	17.3	57	16.8	97	17.1	137	16.7	177	16.9
18	17.2	58	17.6	98	17.1	138	16.8	178	17.1
19	17.4	59	17.2	99	16.7	139	16.7	179	17.2
20	16.8	60	16.6	100	16.9	140	16.7	180	17.4
21	17.1	61	17.1	101	16.5	141	16.6	181	17.5
22	17.4	62	16.9	102	17.2	142	16.5	182	17.9
23	17.4	63	16.6	103	16.4	143	17.0	183	17.0
24	17.5	64	18.0	104	17.0	144	16.7	184	17.0
25	17.4	65	17.2	105	17.0	145	16.7	185	17.0
26	17.6	66	17.3	106	16.7	146	16.9	186	17.2
27	17.4	67	17.0	107	16.2	147	17.4	187	17.3
28	17.3	68	16.9	108	16.6	148	17.1	188	17.4
29	17.0	69	17.3	109	16.9	149	17.0	189	17.4
30	17.8	70	16.8	110	16.5	150	16.8	190	17.0
31	17.5	71	17.3	111	16.6	151	17.2	191	18.0
32	18.1	72	17.4	112	16.6	152	17.2	192	18.2
33	17.5	73	17.7	113	17.0	153	17.4	193	17.6
34	17.4	74	16.8	114	17.1	154	17.2	194	17.8
35	17.4	75	16.9	115	17.1	155	16.9	195	17.7
36	17.1	76	17.0	116	16.7	156	16.8	196	17.2
37	17.6	77	16.9	117	16.8	157	17.0	197	17.4
38	17.7	78	17.0	118	16.3	158	17.4		
39	17.4	79	16.6	119	16.6	159	17.2		
40	17.8	80	16.7	120	16.8	160	17.2		

* 197 个观测值.

序列 B　IBM 普通股收盘价 *

(1961.5.17—1962.11.2　日值)

460	465	545	595	564	552	453	375	382
457	466	549	598	567	551	446	379	384
452	467	547	598	561	550	455	376	382
459	471	543	595	559	553	452	382	383
462	471	540	595	553	554	457	370	383
459	467	539	592	553	551	449	365	388
463	473	532	588	553	551	450	367	395
479	481	517	582	547	545	435	372	392
493	488	527	576	550	547	415	373	386
490	490	540	578	544	547	398	363	383
492	489	542	589	541	537	399	371	377
498	489	538	585	532	539	361	369	364
499	485	541	580	525	538	383	376	369
497	491	541	579	542	533	393	387	355
496	492	547	584	555	525	385	387	350
490	494	553	581	558	513	360	376	353
489	499	539	581	551	510	364	385	340
478	489	557	577	551	521	365	385	350
487	500	557	577	552	521	370	380	349
491	497	560	578	553	521	374	373	358
487	494	571	580	557	523	359	382	360
482	495	571	586	557	516	335	377	360
479	500	569	583	548	511	323	376	366
478	504	575	581	547	518	306	379	359
479	513	580	576	545	517	333	386	356
477	511	584	571	545	520	330	387	355
479	514	585	575	539	519	336	386	367
475	510	590	575	539	519	328	389	357
479	509	599	573	535	519	316	394	361
476	515	603	577	537	518	320	393	355
476	519	599	582	535	513	332	409	348
478	523	596	584	536	499	320	411	343
479	519	585	579	537	485	333	409	330
477	523	587	572	543	454	344	408	340
476	531	585	577	548	462	339	393	339
475	547	581	571	546	473	350	391	331
475	551	583	560	547	482	351	388	345
473	547	592	549	548	486	350	396	352
474	541	592	556	549	475	345	387	346
474	545	596	557	553	459	350	383	352
474	549	596	563	553	451	359	388	357

* 369 个观测值 (按列依次读取).

序列 B′ IBM 普通股收盘价 *

(1959.6.29—1960.6.30, 日值)

445	426	408	439	422	445	499
448	425	415	437	422	439	485
450	424	413	439	417	443	491
447	425	413	438	420	445	496
451	425	410	435	417	450	504
453	424	405	433	418	461	504
454	425	410	437	419	471	409
454	421	412	437	419	467	511
459	414	413	444	417	462	524
440	410	411	441	419	456	525
446	411	411	440	422	464	541
443	406	409	441	423	463	531
443	406	406	439	422	465	529
440	413	407	439	421	464	530
439	411	410	438	421	456	531
435	410	408	437	419	460	527
435	405	408	441	418	458	525
436	409	409	442	421	453	519
435	410	410	441	420	453	514
435	405	409	437	413	449	509
435	401	405	427	413	447	505
433	401	406	423	408	453	513
429	401	405	424	409	450	525
428	414	407	428	415	459	519
425	419	409	428	415	457	519
427	425	407	431	420	453	522
425	423	409	425	420	455	522
422	411	425	423	424	453	
409	414	425	420	426	450	
407	420	428	426	423	456	
423	412	436	418	423	461	
422	415	442	416	425	463	
417	412	442	419	431	463	
421	412	433	418	436	461	
424	411	435	416	436	465	
414	412	433	419	440	473	
419	409	435	425	436	473	
429	407	429	421	443	475	

* 369 个观测值 (按列依次读取).

序列 C 化工生产过程的温度读数 *

(每分钟)

26.6	18.9	25.8	23.8	21.9	23.8
27.0	18.9	25.6	23.7	22.2	23.7
27.1	19.2	25.4	23.6	22.5	23.7
27.1	19.3	25.2	23.4	22.8	23.6
27.1	19.3	24.9	23.2	23.1	23.7
27.1	19.4	24.7	23.0	23.4	23.6
26.9	19.5	24.5	22.8	23.8	23.6
26.8	19.6	24.4	22.6	24.1	23.6
26.7	19.6	24.4	22.4	24.6	23.5
26.4	19.6	24.4	22.0	24.9	23.5
26.0	19.6	24.4	21.6	24.9	23.4
25.8	19.6	24.4	21.3	25.1	23.3
25.6	19.7	24.3	21.2	25.0	23.3
25.2	19.9	24.4	21.2	25.0	23.3
25.0	20.0	24.4	21.1	25.0	23.4
24.6	20.1	24.4	21.0	25.0	23.4
24.2	20.2	24.4	20.9	24.9	23.3
24.0	20.3	24.4	21.0	24.8	23.2
23.7	20.6	24.5	21.0	24.7	23.3
23.4	21.6	24.5	21.1	24.6	23.3
23.1	21.9	24.4	21.2	24.5	23.2
22.9	21.7	24.3	21.1	24.5	23.1
22.8	21.3	24.2	20.9	24.5	22.9
22.7	21.2	24.2	20.8	24.5	22.8
22.6	21.4	24.0	20.8	24.5	22.6
22.4	21.7	23.9	20.8	24.5	22.4
22.2	22.2	23.7	20.8	24.5	22.2
22.0	23.0	23.6	20.9	24.4	21.8
21.8	23.8	23.5	20.8	24.4	21.3
21.4	24.6	23.5	20.8	24.2	20.8
20.9	25.1	23.5	20.7	24.2	20.2
20.3	25.6	23.5	20.7	24.1	19.7
19.7	25.8	23.5	20.8	24.1	19.3
19.4	26.1	23.7	20.9	24.0	19.1
19.3	26.3	23.8	21.2	24.0	19.0
19.2	26.3	23.8	21.4	24.0	18.8
19.1	26.2	23.9	21.7	23.9	
19.0	26.0	23.9	21.8	23.8	

* 369 个观测值 (按列依次读取).

序列 D 化工生产过程的粘度读数 *

(每分钟)

8.0	9.0	9.1	9.5	9.2	9.6	9.0	9.0
8.0	8.8	9.1	9.5	9.2	9.8	9.4	9.0
7.4	8.6	8.9	9.5	9.6	10.2	9.4	9.4
8.0	8.6	9.3	9.9	9.6	10.0	9.4	10.0
8.0	8.8	9.1	9.9	9.8	10.0	9.6	10.0
8.0	8.8	9.1	9.5	9.8	10.0	9.4	10.0
8.0	8.6	9.3	9.3	10.0	9.4	9.6	10.2
8.8	8.6	9.5	9.5	10.0	9.2	9.6	10.0
8.4	8.4	9.3	9.5	9.4	9.6	9.8	10.0
8.4	8.3	9.3	9.1	9.8	9.7	9.8	9.6
8.0	8.4	9.3	9.3	8.8	9.7	9.8	9.0
8.2	8.3	9.9	9.5	8.8	9.8	9.6	9.0
8.2	8.3	9.7	9.3	8.8	9.8	9.2	8.6
8.2	8.1	9.1	9.1	8.8	9.8	9.6	9.0
8.4	8.2	9.3	9.3	9.6	10.0	9.2	9.6
8.4	8.3	9.5	9.1	9.6	10.0	9.2	9.6
8.4	8.5	9.4	9.5	9.6	8.6	9.6	9.0
8.6	8.1	9.0	9.4	9.2	9.0	9.6	9.0
8.8	8.1	9.0	9.5	9.2	9.4	9.6	8.9
8.6	7.9	8.8	9.6	9.0	9.4	9.6	8.8
8.6	8.3	9.0	10.2	9.0	9.4	9.6	8.7
8.6	8.1	8.8	9.8	9.0	9.4	9.6	8.6
8.6	8.1	8.6	9.6	9.4	9.4	10.0	8.3
8.6	8.1	8.6	9.6	9.0	9.6	10.0	7.9
8.8	8.4	8.0	9.4	9.0	10.0	10.4	8.5
8.9	8.7	8.0	9.4	9.4	10.0	10.4	8.7
9.1	9.0	8.0	9.4	9.4	9.8	9.8	8.9
9.5	9.3	8.0	9.4	9.6	9.8	9.0	9.1
8.5	9.3	8.6	9.6	9.4	9.7	9.6	9.1
8.4	9.5	8.0	9.6	9.6	9.6	9.8	9.1
8.3	9.3	8.0	9.4	9.6	9.4	9.6	
8.2	9.5	8.0	9.4	9.6	9.2	8.6	
8.1	9.5	7.6	9.0	10.0	9.0	8.0	
8.3	9.5	8.6	9.4	10.0	9.4	8.0	
8.4	9.5	9.6	9.4	9.6	9.6	8.0	
8.7	9.5	9.6	9.6	9.2	9.6	8.0	
8.8	9.5	10.0	9.4	9.2	9.6	8.4	
8.8	9.9	9.4	9.2	9.2	9.6	8.8	
9.2	9.5	9.3	8.8	9.0	9.6	8.4	
9.6	9.7	9.2	8.8	9.0	9.6	8.4	

* 369 个观测值 (按列依次读取).

序列 E　Wölfer 太阳黑子数 *

(年度值)

1770	101	1795	21	1820	16	1845	40
1771	82	1796	16	1821	7	1846	62
1772	66	1797	6	1822	4	1847	98
1773	35	1798	4	1823	2	1848	124
1774	31	1799	7	1824	8	1849	96
1775	7	1800	14	1825	17	1850	66
1776	20	1801	34	1826	36	1851	64
1777	92	1802	45	1827	50	1852	54
1778	154	1803	43	1828	62	1853	39
1779	125	1804	48	1829	67	1854	21
1780	85	1805	42	1830	71	1855	7
1781	68	1806	28	1831	48	1856	4
1782	38	1807	10	1832	28	1857	23
1783	23	1808	8	1833	8	1858	55
1784	10	1809	2	1834	13	1859	94
1785	24	1810	0	1835	57	1860	96
1786	83	1811	1	1836	122	1861	77
1787	132	1812	5	1837	138	1862	59
1788	131	1813	12	1838	103	1863	44
1789	118	1814	14	1839	86	1864	47
1790	90	1815	35	1840	63	1865	30
1791	67	1816	46	1841	37	1866	16
1792	60	1817	41	1842	24	1867	7
1793	47	1818	30	1843	11	1868	37
1794	41	1819	24	1844	15	1869	74

* 100 个观测值

序列 F 化工生产过程的产量 *

47	71	51	50	48	38	68
64	35	57	71	55	59	38
23	57	50	56	45	55	50
71	40	60	74	57	41	60
38	58	45	50	50	53	39
64	44	57	58	62	49	59
55	80	50	45	44	34	40
41	55	45	54	64	35	57
59	37	25	36	43	54	54
48	74	59	54	52	45	23

* 70 个观测值 (按列依次读取), 该序列曾在表 2.1 中出现.

序列 G 国际航线月度旅客总数

(1949.01—1960.12, 单位: 千人)

	1949	1959	1951	1952	1953	1954	1959	1956	1957	1958	1959	1960
1 月	112	115	145	171	196	204	242	284	315	340	360	417
2 月	118	126	150	180	196	188	233	277	301	318	342	391
3 月	132	141	178	193	236	235	267	317	356	362	406	419
4 月	129	135	163	181	235	227	269	313	348	348	396	461
5 月	121	125	172	183	229	234	270	318	355	363	420	472
6 月	135	149	178	218	243	264	315	374	422	435	472	535
7 月	148	170	199	230	264	302	364	413	465	491	548	622
8 月	148	170	199	242	272	293	347	405	467	505	559	606
9 月	136	158	184	209	237	259	312	355	404	404	463	508
10 月	119	133	162	191	211	229	274	306	347	359	407	461
11 月	104	114	146	172	180	203	237	271	305	310	362	390
12 月	118	140	166	194	201	229	278	306	336	337	405	432

* 144 个观测值.

序列 J　煤气炉数据 *

t	X_t	Y_t	t	X_t	Y_t	t	X_t	Y_t
1	−0.109	53.8	49	1.214	45.6	97	1.054	51.6
2	0.000	53.6	50	1.239	46.0	98	0.595	51.0
3	0.178	53.5	51	1.608	46.9	99	−0.080	50.5
4	0.339	53.5	52	1.905	47.8	100	−0.314	50.4
5	0.373	53.4	53	2.023	48.2	101	−0.288	51.0
6	0.441	53.1	54	1.815	48.3	102	−0.135	51.8
7	0.461	52.7	55	0.535	47.9	103	−0.109	52.4
8	0.348	52.4	56	0.122	47.2	104	−0.187	53.0
9	0.127	52.2	57	0.009	47.2	105	−0.255	53.4
10	−0.180	52.0	58	0.164	48.1	106	−0.229	53.6
11	−0.588	52.0	59	0.671	49.4	107	−0.007	53.7
12	−1.055	52.4	60	1.019	50.6	108	0.254	53.8
13	−1.421	53.0	61	1.146	51.5	109	0.330	53.8
14	−1.520	54.0	62	1.115	51.6	110	0.102	53.8
15	−1.302	54.9	63	1.112	51.2	111	−0.423	53.3
16	−0.814	56.0	64	1.121	50.5	112	−1.139	53.0
17	−0.475	56.8	65	1.223	50.1	113	−2.275	52.9
18	−0.193	56.8	66	1.257	49.8	114	−2.594	53.4
19	0.008	56.4	67	1.157	49.6	115	−2.716	54.6
20	0.435	55.7	68	0.913	49.4	116	−2.510	56.4
21	0.771	55.0	69	0.620	49.3	117	−1.790	58.0
22	0.866	54.3	70	0.255	49.2	118	−1.346	59.4
23	0.875	53.2	71	−0.280	49.3	119	−1.081	60.2
24	0.891	52.3	72	−1.080	49.7	120	−0.910	60.0
25	0.987	51.6	73	−1.551	50.3	121	−0.876	59.4
26	1.263	51.2	74	−1.799	51.3	122	−0.885	58.4
27	1.775	50.8	75	−1.825	52.8	123	−0.800	57.6
28	1.976	50.5	76	−1.456	54.4	124	−0.544	56.9
29	1.934	50.0	77	−0.944	56.0	125	−0.416	56.4
30	1.866	49.2	78	−0.570	56.9	126	−0.271	56.0
31	1.832	48.7	79	−0.431	57.5	127	0.000	55.7
32	1.767	47.9	80	−0.577	57.3	128	0.403	55.3
33	1.608	47.6	81	−0.960	56.6	129	0.841	55.0
34	1.265	47.5	82	−1.616	56.0	130	1.285	54.4
35	0.790	47.5	83	−1.875	55.4	131	1.607	53.7
36	0.360	47.6	84	−1.891	55.4	132	1.746	52.8
37	0.115	48.1	85	−1.746	56.4	133	1.683	51.6
38	0.088	49.0	86	−1.474	57.2	134	1.485	50.6
39	0.331	50.0	87	−1.201	58.0	135	0.993	49.4
40	0.645	51.1	88	−0.927	58.4	136	0.648	48.8
41	0.960	51.8	89	−0.524	58.4	137	0.577	48.5
42	1.409	51.9	90	0.040	58.1	138	0.577	48.7
43	2.670	51.7	91	0.788	57.7	139	0.632	49.2
44	2.834	51.2	92	0.943	57.0	140	0.747	49.8
45	2.812	50.0	93	0.930	56.0	141	0.900	50.4
46	2.483	48.3	94	1.006	54.7	142	0.993	50.7
47	1.929	47.0	95	1.137	53.2	143	0.968	50.9
48	1.485	45.8	96	1.198	52.1	144	0.790	50.7

* 每隔 9 秒取样, 共 296 对观测值, X,0.60–0.40, 输入煤气速度 (立方米 / 秒);Y, 排出煤气的 CO_2 比例.

序列 J 煤气炉数据 (续)

t	X_t	Y_t
145	0.399	50.5
146	−0.161	50.4
147	−0.553	50.2
148	−0.603	50.4
149	−0.424	51.2
150	−0.194	52.3
151	−0.049	53.2
152	0.060	53.9
153	0.161	54.1
154	0.301	54.0
155	0.517	53.6
156	0.566	53.2
157	0.560	53.0
158	0.573	52.8
159	0.592	52.3
160	0.671	51.9
161	0.933	51.6
162	1.337	51.6
163	1.460	51.4
164	1.353	51.2
165	0.772	50.7
166	0.218	50.0
167	−0.237	49.4
168	−0.714	49.3
169	−1.099	49.7
170	−1.269	50.6
171	−1.175	51.8
172	−0.676	53.0
173	0.033	54.0
174	0.556	55.3
175	0.643	55.9
176	0.484	55.9
177	0.109	54.6
178	−0.310	53.5
179	−0.697	52.4
180	−1.047	52.1
181	−1.218	52.3
182	−1.183	53.0
183	−0.873	53.8
184	−0.336	54.6
185	0.063	55.4
186	0.084	55.9
187	0.000	55.9
188	0.001	55.2
189	0.209	54.4
190	0.556	53.7
191	0.782	53.6
192	0.858	53.6
193	0.918	53.2
194	0.862	52.5
195	0.416	52.0
196	−0.336	51.4
197	−0.959	51.0
198	−1.813	50.9
199	−2.378	52.4
200	−2.499	53.5
201	−2.473	55.6
202	−2.330	58.0
203	−2.053	59.5
204	−1.739	60.0
205	−1.261	60.4
206	−0.569	60.5
207	−0.137	60.2
208	−0.024	59.7
209	−0.050	59.0
210	−0.135	57.6
211	−0.276	56.4
212	−0.534	55.2
213	−0.871	54.5
214	−1.243	54.1
215	−1.439	54.1
216	−1.422	54.4
217	−1.175	55.5
218	−0.813	56.2
219	−0.634	57.0
220	−0.582	57.3
221	−0.625	57.4
222	−0.713	57.0
223	−0.848	56.4
224	−1.039	55.9
225	−1.346	55.5
226	−1.628	55.3
227	−1.619	55.2
228	−1.149	55.4
229	−0.488	56.0
230	−0.160	56.5
231	−0.007	57.1
232	−0.092	57.3
233	−0.620	56.8
234	−1.086	55.6
235	−1.525	55.0
236	−1.858	54.1
237	−2.029	54.3
238	−2.024	55.3
239	−1.961	56.4
240	−1.952	57.2
241	−1.794	57.8
242	−1.302	58.3
243	−1.030	58.6
244	−0.918	58.8
245	−0.798	58.8
246	−0.867	58.6
247	−1.047	58.0
248	−1.123	57.4
249	−0.876	57.0
250	−0.395	56.4
251	0.185	56.3
252	0.662	56.4
253	0.709	56.4
254	0.605	56.0
255	0.501	55.2
256	0.603	54.0
257	0.943	53.0
258	1.223	52.0
259	1.249	51.6
260	0.824	51.6
261	0.102	51.1
262	0.025	50.4
263	0.382	50.0
264	0.922	50.0
265	1.032	52.0
266	0.866	54.0
267	0.527	55.1
268	0.093	54.5
269	−0.458	52.8
270	−0.748	51.4
271	−0.947	50.8
272	−1.029	51.2
273	−0.928	52.0
274	−0.645	52.8
275	−0.424	53.8
276	−0.276	54.5
277	−0.158	54.9
278	−0.033	54.9
279	0.102	54.8
280	0.251	54.4
281	0.280	53.7
282	0.000	53.3
283	−0.493	52.8
284	−0.759	52.6
285	−0.824	52.6
286	−0.740	53.0
287	−0.528	54.3
288	−0.204	56.0
289	0.034	57.0
290	0.204	58.0
291	0.253	58.6
292	0.195	58.5
293	0.131	58.3
294	0.017	57.8
295	−0.182	57.3
296	−0.262	57.0

序列 K　两输入的动态模拟数据 *

t	X_{1t}	X_{2t}	Y_t	t	X_{1t}	X_{2t}	Y_t
−2			58.3	30			65.8
−1	0	0	61.8	31			67.4
0			64.2	32	−1	−1	64.7
1			62.1	33			65.7
2	−1	1	55.1	34			67.5
3			50.6	35			58.2
4			47.8	36			57.0
5			49.7	37	−1	1	54.7
6			51.6	38			54.9
7	1	−1	58.5	39			48.4
8			61.5	40			49.7
9			63.3	41			53.1
10			65.9	42	1	−1	50.2
11			70.9	43			51.7
12	−1	−1	65.8	44			57.4
13			57.6	45			62.6
14			56.1	46			65.8
15			58.2	47	−1	−1	61.5
16			61.7	48			61.5
17	1	1	59.2	49			56.8
18			57.9	50			62.3
19			61.3	51			57.7
20			60.8	52	−1	1	54.0
21			63.6	53			45.2
22	1	−1	69.5	54			51.9
23			69.3	55			45.6
24			70.5	56			46.2
25			68.0	57	1	1	50.2
26			68.1	58			54.6
27	1	1	65.0	59			55.6
28			71.9	60	0	0	60.4
29			64.8	61			59.4

* 64 对观测值.

序列 L 试验方案数据 *

t	x_t	ϵ_t	t	x_t	ϵ_t	t	x_t	ϵ_t	t	x_t	ϵ_t
1	30	−4	41	0	0	81	−100	12	121	10	−4
2	0	−2	42	0	0	82	0	8	122	20	−4
3	−10	0	43	20	−2	83	0	−12	123	10	−3
4	0	0	44	−50	4	84	50	−15	124	0	−2
5	−40	4	45	20	0	85	85	−15	125	−70	6
6	0	2	46	0	0	86	5	−12	126	50	−2
7	−10	2	47	0	0	87	40	−14	127	30	−4
8	10	0	48	40	−4	88	10	−8	128	0	−2
9	20	−2	49	0	−2	89	−60	2	129	−10	0
10	50	−6	50	50	−6	90	−50	6	130	0	0
11	−10	−2	51	−40	0	91	−50	8	131	−40	4
12	−55	4	52	−50	3	92	40	0	132	0	2
13	0	2	53	−60	6	93	0	0	133	−10	2
14	10	0	54	50	−2	94	0	0	134	10	0
15	0	−2	55	−10	0	95	−20	2	135	0	0
16	10	−2	56	40	−4	96	−30	4	136	80	−8
17	−70	6	57	40	−6	97	−60	8	137	−80	4
18	30	0	58	−30	0	98	−20	6	138	20	4
19	−20	2	59	20	−2	99	−30	6	139	20	0
20	10	0	60	−30	2	100	30	0	140	−10	2
21	0	0	61	10	0	101	−40	4	141	10	0
22	0	0	62	−20	2	102	80	−6	142	0	0
23	20	−2	63	30	−2	103	−40	0	143	−20	2
24	30	−4	64	−50	4	104	−20	2	144	20	−1
25	0	−2	65	10	−2	105	55	−4	145	55	−6
26	−10	0	66	10	−2	106	0	2	146	0	−3
27	−20	2	67	10	−2	107	−90	8	147	25	−4
28	−30	4	68	−30	0	108	40	0	148	20	−4
29	0	2	69	0	0	109	0	0	149	−60	4
30	10	0	70	−10	2	110	80	−8	150	−40	6
31	20	−2	71	−10	3	111	−20	−2	151	10	4
32	−10	0	72	15	0	112	−10	0	152	20	0
33	0	0	73	20	−2	113	−70	6	153	60	−6
34	20	−2	74	−50	4	114	−30	6	154	−50	2
35	10	−2	75	20	0	115	−10	4	155	−10	2
36	−10	0	76	0	0	116	30	−1	156	−30	4
37	0	0	77	0	0	117	−5	0	157	20	0
38	0	0	78	0	0	118	−60	6	158	0	0
39	0	0	79	0	0	119	70	−4	159	20	−2
40	0	0	80	−40	4	120	40	−6	160	10	−2

* 312 对观测值.

序列 L 试验方案数据 (续)

t	x_t	ϵ_t	t	x_t	ϵ_t	t	x_t	ϵ_t	t	x_t	ϵ_t
161	10	−2	201	−75	8	241	−85	8	281	−110	13
162	10	−22	202	−40	8	242	0	4	282	15	4
163	50	−6	203	−40	6	243	0	0	283	30	−2
164	−30	0	204	90	−6	244	60	−4	284	0	−1
165	−30	6	205	90	−12	245	40	−6	285	25	−3
166	−90	12	206	80	−14	246	−30	0	286	− 5	−1
167	60	0	207	−45	−2	247	−40	4	287	−15	1
168	−40	4	208	−10	0	248	−40	6	288	45	−4
169	20	0	209	−40	4	249	50	−2	289	40	−6
170	0	0	210	40	−2	250	10	−2	290	−50	2
171	20	−2	211	−90	8	251	30	−4	291	−10	2
172	10	−2	212	40	0	252	−40	2	292	−50	6
173	−30	2	213	0	0	253	10	0	293	20	1
174	−30	4	214	0	0	254	−40	4	294	5	0
175	0	2	215	0	0	255	40	−2	295	−40	4
176	50	−4	216	20	−2	256	−30	2	296	0	6
177	−60	4	217	90	−10	257	−50	6	297	−60	8
178	20	0	218	30	−8	258	0	3	298	40	0
179	0	0	219	20	−6	259	−45	6	299	−20	2
180	40	−8	220	30	−6	260	−20	5	300	130	−12
181	80	−1	221	30	−6	261	−25	4	301	−20	−4
182	20	−8	222	30	−6	262	35	−2	302	0	−2
183	−100	6	223	30	−6	263	70	8	303	30	−4
184	−30	6	224	−90	6	264	−10	−5	304	−20	0
185	30	0	225	10	2	265	100	−20	305	60	6
186	−20	2	226	10	2	266	−20	−8	306	10	−4
187	−30	4	227	−30	4	267	−40	0	307	−10	1
188	20	0	228	−20	4	268	−20	2	308	−25	2
189	60	−6	229	40	−2	269	10	0	309	0	1
190	−10	−2	230	10	−2	270	0	0	310	15	−1
191	30	−4	231	10	−2	271	0	0	311	− 5	0
192	−40	2	232	10	−2	272	−20	2	312	0	0
193	30	−2	233	−100	12	273	−50	6			
194	−20	1	234	10	6	274	50	−2			
195	5	0	235	45	−2	275	30	−4			
196	−20	2	236	30	−4	276	60	−8			
197	−30	4	237	30	−5	277	−40	0			
198	20	0	238	−15	−1	278	−20	2			
199	10	−1	239	− 5	0	279	−10	2			
200	−15	1	240	10	−1	280	10	0			

序列 M　带有领先指标的销售额数据 *

	领先指标	销售额		领先指标	销售额		领先指标	销售额		领先指标	销售额
t	X_t	Y_t	t	X_t	Y_t	t	X_t	Y_t	t	X_t	Y_t
1	10.01	200.1	41	11.01	215.3	81	11.28	212.5	121	13.15	261.6
2	10.07	199.5	42	11.22	215.9	82	11.27	214.8	122	13.30	260.8
3	10.32	199.4	43	11.21	216.7	83	11.44	215.3	123	12.94	259.8
4	9.75	198.9	44	11.91	216.7	84	11.52	217.5	124	13.29	259.0
5	10.33	199.0	45	11.69	217.7	85	12.10	218.8	125	13.26	258.9
6	10.13	200.2	46	10.93	218.7	86	11.83	220.7	126	13.08	257.4
7	10.36	198.6	47	10.99	222.9	87	12.62	222.2	127	13.24	257.7
8	10.32	200.0	48	11.01	224.9	88	12.41	226.7	128	13.31	257.9
9	10.13	200.3	49	10.84	222.2	88	12.43	228.4	129	13.52	257.4
10	10.16	201.2	50	10.76	220.7	90	12.73	233.2	130	13.02	257.3
11	10.58	201.6	51	10.77	220.0	91	13.01	235.7	131	13.25	257.6
12	10.62	201.5	52	10.88	218.7	92	12.74	237.1	132	13.12	258.9
13	10.86	201.5	53	10.49	217.0	93	12.73	240.6	133	13.26	257.8
14	11.20	203.5	54	10.50	215.9	94	12.76	243.8	134	13.11	257.7
15	10.74	204.9	55	11.00	215.8	95	12.92	245.3	135	13.30	257.2
16	10.56	207.1	56	10.98	214.1	96	12.64	246.0	136	13.06	257.5
17	10.48	210.5	57	10.61	212.3	97	12.79	246.3	137	13.32	256.8
18	10.77	210.5	58	10.48	213.9	98	13.05	247.7	138	13.10	257.5
19	11.33	209.8	59	10.53	214.6	99	12.69	247.6	139	13.27	257.0
20	10.96	208.8	60	11.07	213.6	100	13.01	247.8	140	13.64	257.6
21	11.16	209.5	61	10.61	212.1	101	12.90	249.4	141	13.58	257.3
22	11.70	213.2	62	10.86	211.4	102	13.12	249.0	142	13.87	257.5
23	11.39	213.7	63	10.34	213.1	103	12.47	249.9	143	13.53	259.6
24	11.42	215.1	64	10.78	212.9	104	12.47	250.5	144	13.41	261.1
25	11.94	218.7	65	10.80	213.3	105	12.94	251.5	145	13.25	262.9
26	11.24	219.8	66	10.33	211.5	106	13.10	249.0	146	13.50	263.3
27	11.59	220.5	67	10.44	212.3	107	12.91	247.6	147	13.58	262.8
28	10.96	223.8	68	10.50	213.0	108	13.39	248.8	148	13.51	261.8
29	11.40	222.8	69	10.75	211.0	109	13.13	250.4	149	13.77	262.2
30	11.02	223.8	70	10.40	210.7	110	13.34	250.7	150	13.40	262.7
31	11.01	221.7	71	10.40	210.1	111	13.34	253.0			
32	11.23	222.3	72	10.34	211.4	112	13.14	253.7			
33	11.33	220.8	73	10.55	210.0	113	13.49	255.0			
34	10.83	219.4	74	10.46	209.7	114	13.87	256.2			
35	10.84	220.1	75	10.82	208.8	115	13.39	256.0			
36	11.14	220.6	76	10.91	208.8	116	13.59	257.4			
37	10.38	218.9	77	10.87	208.8	117	13.27	260.4			
38	10.90	217.8	78	10.67	210.6	118	13.70	260.0			
39	11.05	217.7	79	11.11	211.9	119	13.20	261.3			
40	11.11	215.0	80	10.88	212.8	120	13.32	260.4			

* 150 对观测值.

序列 N 哈德逊湾 (Husdon Bay) 公司貂皮销售量数据 *

(1850–1911, 年度值)

年	值	年	值	年	值	年	值
1850	29619	1866	51404	1882	45600	1897	76365
1851	21151	1867	58451	1883	47508	1898	70407
1852	24859	1868	73575	1884	52290	1899	41839
1853	25152	1869	74343	1885	110824	1900	45978
1854	42375	1870	27708	1886	76503	1901	47813
1855	50839	1871	31985	1887	64303	1902	57620
1856	61581	1872	39266	1888	83023	1903	66549
1857	61951	1873	44740	1889	40748	1904	54673
1858	76231	1874	60429	1890	35596	1905	55996
1859	63264	1875	72237	1891	29479	1906	60053
1860	44730	1876	79214	1892	42264	1907	39169
1861	31094	1877	79060	1893	58171	1908	21534
1862	49452	1878	84244	1894	50815	1909	17857
1863	43961	1879	62590	1895	51285	1910	21788
1864	61727	1880	35072	1896	70229	1911	33008
1865	60334	1881	36160				

* 62 个观测值.

序列 P 英国失业和 GDP 数据 *

(1955–1969, 季度值)

		UN	GPT			UN	GPT			UN	GPT
1955	1	225	81.37	1960	1	363	92.30	1965	1	306	108.07
	2	208	82.60		2	342	92.13		2	304	107.64
	3	201	82.30		3	325	93.17		3	321	108.87
	4	199	83.00		4	312	93.50		4	305	109.75
1956	1	207	82.87	1961	1	291	94.77	1966	1	279	110.20
	2	215	83.60		2	293	95.37		2	282	110.20
	3	240	83.33		3	304	95.03		3	318	110.90
	4	245	83.53		4	330	95.23		4	414	110.40
1957	1	295	84.27	1962	1	357	95.07	1967	1	463	111.00
	2	293	85.50		2	401	96.40		2	506	112.10
	3	279	84.33		3	447	96.97		3	538	112.50
	4	287	84.30		4	483	96.50		4	536	113.00
1958	1	331	85.07	1963	1	535	96.16	1968	1	544	114.30
	2	396	83.60		2	520	99.79		2	541	115.10
	3	432	84.37		3	489	101.14		3	547	116.40
	4	462	84.50		4	456	102.95		4	532	117.80
1959	1	454	85.20	1964	1	386	103.96	1969	1	532	116.80
	2	446	87.07		2	368	105.28		2	519	117.80
	3	426	88.40		3	358	105.81		3	547	119.00
	4	402	90.03		4	330	107.14		4	544	119.60

* 60 对观测值, 数据经季节调整, 失业人数 (NU) 按千人计; 国民生产总值是综合估算 (1963=100) . 数据引自 [214] .

序列 Q 美国生猪价格数据 *

(1867–1948, 年度值)

1867	597	1888	709	1909	810	1929	1112
1868	509	1889	763	1910	957	1930	1229
1869	663	1890	681	1911	970	1931	1055
1870	751	1891	627	1912	903	1932	787
1871	739	1892	667	1913	995	1933	624
1872	598	1893	804	1914	1022	1934	612
1873	556	1894	782	1915	998	1935	800
1874	594	1895	707	1916	928	1936	1104
1875	667	1896	653	1917	1073	1937	1075
1876	776	1897	639	1918	1294	1938	1052
1877	754	1898	672	1919	1346	1939	1048
1878	689	1899	669	1920	1301	1940	891
1879	498	1900	729	1921	1134	1941	921
1880	643	1901	784	1922	1024	1942	1193
1881	681	1902	842	1923	1090	1943	1352
1882	778	1903	886	1924	1013	1944	1243
1883	829	1904	784	1925	1119	1945	1314
1884	751	1905	770	1926	1195	1946	1380
1885	704	1906	783	1927	1235	1947	1556
1886	633	1907	877	1928	1120	1948	1632
1887	663	1908	777				

* 82 个观测值, 取值为 $1000\log_{10}(H_t)$ 其中 H_t 是每年 1 月第一天的价格 (美元), 数据引自 [162] .

序列 R 洛杉矶市臭氧每小时读数的月平均值 *

	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1955	2.63	1.94	3.38	4.92	6.29	5.58	5.50	4.71	6.04	7.13	7.79	3.83
1956	3.83	4.25	5.29	3.75	4.67	5.42	6.04	5.71	8.13	4.88	5.42	5.50
1957	3.00	3.42	4.50	4.25	4.00	5.33	5.79	6.58	7.29	5.04	5.04	4.48
1958	3.33	2.88	2.50	3.83	4.17	4.42	4.25	4.08	4.88	4.54	4.25	4.21
1959	2.75	2.42	4.50	5.21	4.00	7.54	7.38	5.96	5.08	5.46	4.79	2.67
1960	1.71	1.92	3.38	3.98	4.63	4.88	5.17	4.83	5.29	3.71	2.46	2.17
1961	2.15	2.44	2.54	3.25	2.81	4.21	4.13	4.17	3.75	3.83	2.42	2.17
1962	2.33	2.00	2.13	4.46	3.17	3.25	4.08	5.42	4.50	4.88	2.83	2.75
1963	1.63	3.04	2.58	2.92	3.29	3.71	4.88	4.63	4.83	3.42	2.38	2.33
1964	1.50	2.25	2.63	2.96	3.46	4.33	5.42	4.79	4.38	4.54	2.04	1.33
1965	2.04	2.81	2.67	4.08	3.90	3.96	4.50	5.58	4.52	5.88	3.67	1.79
1966	1.71	1.92	3.58	4.40	3.79	5.52	5.50	5.00	5.48	4.81	2.42	1.46
1967	1.71	2.46	2.42	1.79	3.63	3.54	4.88	4.96	3.63	5.46	3.08	1.75
1968	2.13	2.58	2.75	3.15	3.46	3.33	4.67	4.13	4.73	3.42	3.08	1.79
1969	1.96	1.63	2.75	3.06	4.31	3.31	3.71	5.25	3.67	3.10	2.25	2.29
1970	1.25	2.25	2.67	3.23	3.58	3.04	3.75	4.54	4.46	2.83	1.63	1.17
1971	1.79	1.92	2.25	2.96	2.38	3.38	3.38	3.21	2.58	2.42	1.58	1.21
1972	1.42	1.96	3.04	2.92	3.58	3.33	4.04	3.92	3.08	2.00	1.58	1.21

* 216 个观测值, 以亿分之一计.

第六部分
习题和问题

本书的这一部分是前面各章的习题和问题。当使用本书作教材时，我们希望这些习题和问题能进一步提高本书的价值，同时也希望它们对自学有所帮助。许多例子都是围绕一些概念的推广以及对其它一些方法的介绍。

习题和问题

2

2.1 下面是对一个化学反应器每分钟所做的温度测量值 z:

200, 202, 208, 204, 204, 207, 207, 204, 202, 199, 201, 198,
200, 202, 203, 205, 207, 211, 204, 206, 203, 203, 201, 198,
200, 206, 207, 206, 200, 203, 203, 200, 200, 195, 202, 204,

(a) 画出该时间序列的图形.

(b) 画出 z_{t+1} 作为 z_t 函数的图.

(c) 画出 z_{t+2} 作为 z_t 函数的图.

在对这些图进行观察之后, 你认为该序列是自相关的吗?

2.2 说明一个平稳随机过程能否具有下列自相关值.

(a) $\rho_1 = 0.80, \rho_2 = 0.55$, 对于 $k > 2, \rho_k = 0$

(b) $\rho_1 = 0.80, \rho_2 = 0.28$, 对于 $k > 2, \rho_k = 0$

2.3 两个随机过程 z_{1t} 和 z_{2t} 具有下列自相关函数:

$$z_{1t}: \quad \gamma_0 = 0.5, \ \gamma_1 = 0.2, \ \gamma_j = 0 (j \geqslant 2)$$

$$z_{2t}: \quad \gamma_0 = 2.30, \ \gamma_1 = -1.43, \ \gamma_2 = 0.30, \ \gamma_j = 0 (j \geqslant 3)$$

计算过程 $z_{3t} = z_{1t} + z_{2t}$ 的自协方差函数, 并验证它是一个有效的平稳过程.

2.4 计算习题 2.1 中序列的 c_0, c_1, c_2, r_1, r_2, 画出 $r_k, k = 0,1,2$ 的图形.

2.5 在假设 $\rho_j = 0 (j > 2)$ 之下

(a) 得出 r_1, r_2, 及 $r_j, j > 2$ 的近似标准差.

(b) 得出 r_4, r_5, 之间的近似相关.

2.6 利用习题 2.1 中的数据, 计算周期为 36,18,12,9,36/5,5 的周期图, 并画出一个方差分析表说明与这些周期相联系的均方和残差均方.

2.7 一个周期为 N 的循环随机过程定义为 $z_t = z_{t+N}$.

(a) 证明 (参见 [63],[93],[122]) 当 $N = 2n$ 时, z_t 的 $N \times N$ 自相关阵的特征值为

$$\lambda_k = 1 + 2\sum_{i=1}^{n-1} \rho_i \cos\left(\frac{\pi i k}{n}\right) + \rho \cos(\pi k)$$

其中 $k=1,2,\cdots,N$. 并证明对应于 λ_k,λ_{N-k}(有 $\lambda_k=\lambda_{N-k}$) 是

$$l'_k=\left(\cos\left(\frac{\pi k}{n}\right),\cos\left(\frac{2\pi k}{n}\right),\cdots,\cos(2\pi k)\right)$$
$$l'_{N-k}=\left(\sin\left(\frac{\pi k}{n}\right),\sin\left(\frac{2\pi k}{n}\right),\cdots,\sin(2\pi k)\right)$$

(b) 验证：当 N 趋于无穷且 k/N 固定时，λ_k 趋于 $g(k/N)/2$, 这里 $g(f)$ 是谱密度函数，也就是在极限情形，给出自协方差阵特征根描绘的谱曲线.

3

3.1 将下列模型用记号 B 写出:

(1) $\tilde{z}_t-0.5\tilde{z}_{t-1}=a_t$

(2) $\tilde{z}_t=a_t-1.3a_{t-1}+0.4a_{t-2}$

(3) $\tilde{z}_t-0.5\tilde{z}_{t-1}=a_t-1.3a_{t-1}+0.4a_{t-2}$

3.2 对习题 3.1 中的各个模型，得出

(a) 前 4 个 ψ 权.

(b) 前 4 个 π 权.

(c) 自协方差生成函数.

(d) 前 4 个自相关.

(e) 假设 $\sigma_a^2=1.0$ 时 $\tilde{z}_t$ 的方差.

3.3 对于习题 3.1 及下列模型中的每一个说明:

(a) 是否平稳;　(b) 是否可逆.

(4) $\tilde{z}_t-1.5\tilde{z}_{t-1}+0.6\tilde{z}_{t-2}=a_t$

(5) $\tilde{z}_t-\tilde{z}_{t-1}=a_t-0.5a_{t-1}$

(6) $\tilde{z}_t-\tilde{z}_{t-1}=a_t-1.3a_{t-1}+0.3a_{t-2}$

3.4 将习题 3.1 和 3.3 中模型 (1)-(4) 的每一个按 ARMA(p,q) 过程分类.

3.5 (a) 写出模型 (1) 和 (4) 的 Yule-Walker 方程.

(b) 对此二模型求解方程组，并得出 ρ_1 和 ρ_2 .

(c) 对此二模型得出偏自相关函数.

3.6 对于 AR(2) 过程 $\tilde{z}_t-1.0\tilde{z}_{t-1}+0.5\tilde{z}_{t-2}=a_t$,

(a) 计算 ρ_1 .

(b) 以 ρ_0 和 ρ_1 作为初始值，利用自相关函数的差分方程形式计算 $\rho_k,k=2,\cdots,15$.

(c) 利用自相关函数的图形去估计其周期及衰减因子.

(d) 利用 ϕ_1 和 ϕ_2, 通过直接计算来检验 (c) 中诸值.

3.7 (a) 画出习题 3.6 中自回归过程的功率谱 $g(f)$, 并说明它在靠近自相关函数的周期值附近存在一个峰值.

(b) 用图形或其它方法去估计在 f =0.0 和 f =0.2(每个数据间隔上的周数) 之间的频带上, 序列方差的比例.

3.8 (a) 在对一个观测序列进行模型拟合之后, 为什么将自回归算子和滑动平均算子分解因式是很重要的?

(b) 在 [120] 中说明了, 从 1848 年到 1909 年之间加拿大北部年度貂皮交易数量 z_t 可以适用于下面的 AR(4) 模型

$$(1-0.82B+0.22B^2+0.28B^4)[\ln(z_t)-\mu]=a$$

将该自回归算子分解因式, 并说明这些因式对于自相关函数和貂皮序列的基本性质揭示出了什么。上述数据作为序列 N 在本书第五部分的时间序列汇集中给出。

4

4.1 对于下列模型中的每一个

(1) $(1-B)z_t=(1-0.5B)a_t$

(2) $(1-B)z_t=(1-0.2B)a_t$

(3) $(1-0.5B)(1-B)z_t=a_t$

(4) $(1-0.2B)(1-B)z_t=a_t$

(5) $(1-0.2B)(1-B)z_t=(1-0.5B)a_t$

(a) 求出前 7 个 ψ 权.

(b) 求出前 7 个 π 权.

(c) 按 ARIMA(p,d,q) 过程进行分类。

4.2 利用已得到的适当结果, 对于习题 4.1 中的 5 个模型

(a) 将每个模型写成随机冲击形式.

(b) 将每个模型写成余函数加上相对于原点 $k=t-3$ 的特解形式.

(c) 将每个模型写成逆转形式.

4.3 设 z_0 =20,z_{-1} =19, 且随机冲击序列给出如下:

t	a_t	t	a_t	t	a_t
0	−0.3	5	−0.6	10	−0.4
1	0.6	6	1.7	11	0.9
2	0.9	7	−0.9	12	0.0
3	0.2	8	−1.3	13	−1.4
4	0.1	9	−0.6	14	−0.6

(a) 利用模型的差分方程形式，对习题 4.1 的 5 个模型中的每一个求出 $z_1,\cdots,z_{14}$.

(b) 画出导出序列 $\{z_t\}$.

4.4 利用习题 4.1 中每个模型的逆转形式，并且只用习题 4.3 的导出值 $z_1,z_2,\cdots,z_{11}$ 和 a_{12},a_{13},a_{14} 来求出 z_{12},z_{13},z_{14} . 证实这些值与习题 4.3 中所得到的值是一致的.

4.5 若 $\bar{z}_t=\sum\limits_{j=1}^{\infty}\pi_j z_{t-j+1}$, 那么, 对于习题 4.1 中形如 $(1-B)z_t=(1-\theta B)a_t$ 的模型 (1) 和 (2), $\bar{z}_t$ 是指数加权滑动平均. 对于这两个模型通过实际计算验证 $\bar{z}_{11},\bar{z}_{12}$ 和 z_{13} 满足关系式

$$z_t=\bar{z}_{t-1}+a_t \qquad (\text{见习题 } 4.4)$$
$$\bar{z}_t=\bar{z}_{t-1}+(1-\theta)a_t$$
$$\bar{z}_t=(1-\theta)z_t+\theta\bar{z}_{t-1}$$

4.6 如果 $w_{1t}=(1-\theta_1 B)a_{1t}, w_{2t}=(1-\theta_2 B)a_{2t}$, 说明 $w_{3t}=w_{1t}+w_{2t}$ 可以写成 $w_{3t}=(1-\theta_3 B)a_t$. 并推出 θ_3 和 σ_{3a}^2 用前两个过程参数表示的表达式，阐明你的设想.

4.7 假设 $Z_t=z_t+b_t$, 这里 z_t 是一阶自回归过程 $(1-\phi B)z=a_t$, b_t 是方差为 σ_b^2 的白噪声过程， Z_t 服从什么过程，阐明你的设想.

5

5.1 对于下面模型

(1) $\tilde{z}_t-0.5\tilde{z}_{t-1}=a_t$

(2) $\nabla z_t=a_t-0.5a_{t-1}$

(3) $(1-0.6B)\tilde{z}_t=a_t$

表示出提前期为 $l=1$ 和 $l=2$ 的预报，使用

(a) 差分方程

(b) 求和形式 (用 ψ 权)

(c) 过去观测值的加权平均

5.2 用模型 $\nabla z_t = a_t - 1.1a_{t-1} + 0.28a_{t-2}$ 拟合的序列中的 $z_{91}, z_{92}, \cdots, z_{100}$, 其值由下面观测值给出

166, 172, 172, 169, 164, 168, 171, 167, 168, 172

(a) 生成预报值 $\hat{z}_{100}(l), l = 1, 2, \cdots, 12$, 并画出序列值和预报值的图 (假设 $a_{90}=0, a_{91}=0$) .

(b) 利用 $\hat{\sigma}_a^2 = 1.103$, 计算预报误差标准差的估计值 $\hat{\sigma}(l)$, 并由此计算预报的 80% 概率限, 并将概率限加到图中预报值的两侧.

5.3 假设习题 5.2 中的数据表示月度销售额.

(a) 利用直到 $t=100$ 的数据, 对于季度销售额计算提前期为 1, 2, 3, 4 季的最小均方误差预报.

(b) 对这些预报值计算 80% 的概率限.

5.4 由数据和习题 5.2 的预报值, 并给出了未来观测值 $z_{101} = 174$,

(a) 请利用下面修正公式

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1}$$

计算预报值 $\hat{z}_{101}(l), l = 1, 2, \cdots, 11$

(b) 直接利用差分方程验证这些预报.

5.5 对于习题 5.2 中的模型 $\nabla z_t = a_t - 1.1a_{t-1} + 0.28a_{t-2}$

(a) 写出同一原点的预报误差 $e_t(1), e_t(2), \cdots, e_t(6)$ 的表达式.

(b) 计算并画出预报误差序列 $e_t(3)$ 的自相关.

(c) 计算并画出预报误差 $e_t(2)$ 和 $e_t(j), j = 1, 2, \cdots, 6$ 之间的相关.

5.6 向量 $e' = (e_1, e_2, \cdots, e_L)$, 其元素是具有相同原点 t 提前期为 1, 2, $\cdots$, L 的预报误差. 若 $a' = (a_{t+1}, a_{t+2}, \cdots, a_{t+L})$ 是相应的不相关冲击, 且

$$e = Ma \qquad \text{其中} \quad M = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & \cdots & 0 \\ \psi_2 & \psi_1 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \psi_{L-1} & \psi_{L-2} & \psi_{L-3} & \cdots & 1 \end{bmatrix}$$

记 e 的协方差阵为 Σ_e, 证明 (参见 [60],[186]) $\Sigma_e = MM'\sigma_a^2$. 并由此说明, 关于序列的一组顺序实现值 $z_{t+1}, z_{t+2}, \cdots, z_{t+L}$ 与以 t 为原

点的预报值之间从整体上无明显差异的检验，可通过利用

$$e'\Sigma_e^{-1}e=\frac{e'(MM')^{-1}e}{\sigma_a^2}=\frac{a'a}{\sigma_a^2}=\frac{1}{\sigma_a^2}\sum_{j=t+1}^{t+L}a_j^2$$

是自由度为 L 的 χ^2 分布而得出. (注意，a_{t+j} 是由 $z_{t+j}-\hat{z}_{t+j-1}(1)$ 计算的提前 1 期预报误差)．

5.7 对于一个经济时间序列已确知可用模型

$$\nabla z_t=0.5+(1-1.0B+0.5B^2)a_t$$

很好地描述，其中 $\sigma_a^2=0.04$．

(a) 给定 $z_{48}=130, a_{47}=-0.3, a_{48}=0.2$, 计算并画出预报值 $\hat{z}_{48}, l=1,2,\cdots,12$．

(b) 在图中嵌上 80% 概率限.

(c) 用求和形式表示序列和预报值.

6

6.1 下表给出被识别的模型和 $w_t=\nabla^d z_t$ 的自相关估计值:

	识别模型			
	p	d	q	自相关估计值
(1)	1	1	0	$r_1=0.72$
(2)	0	0	1	$r_1=-0.41$
(3)	1	0	1	$r_1=0.40, r_2=0.32$
(4)	0	2	2	$r_1=0.62, r_2=0.13$
(5)	2	1	0	$r_1=0.93, r_2=0.81$

(a) 分析得出参数的初估计.

(b) 利用本书第五部分的表和图检验这些估计.

(c) 代入这些初估计，用后移算子写出被识别模型.

6.2 对于习题 6.1 中 (5) 所考虑的过程 (2,1,0), 设 $w_t=\nabla z_t$ 的样本均值和方差为 $\bar{w}=0.23$ 和 $s_w^2=0.25$．若序列包含 $N=101$ 个观测值.

(a) 说明模型中必须包含一个常数项.

(b) 将模型表示为形式 $w_t-\phi_1 w_{t-1}-\phi_2 w_{t-2}=\theta_0+a_t$, 并用数值代入参数.

6.3 对于英国季度失业人数 (这些数据在 [214] 中给出，其中的一部分列在本书第五部分时间序列汇集的序列 P 中) 经对数化后的 60 个观测值，其 ACFr_k 和 PACF $\hat{\phi}_{kk}$ 的前 16 个值列于下表:

k	r_k	$\hat{\phi}_{kk}$	k	r_k	$\hat{\phi}_{kk}$	k	r_k	$\hat{\phi}_{kk}$
1	0.93	0.93	7	0.03	0.05	13	−0.21	0.19
2	0.80	−0.41	8	−0.09	−0.07	14	−0.12	0.20
3	0.65	−0.14	9	−0.16	0.12	15	−0.01	0.03
4	0.49	−0.11	10	−0.22	−0.14	16	0.10	−0.11
5	0.32	−0.07	11	−0.25	0.03			
6	0.16	−0.10	12	−0.25	0.09			

(a) 画出 ACF 和 PACF 的图形.

(b) 对于该序列识别出一个模型.

(c) 得出参数的初估计及其标准差

(c) 给定 $\bar{z}$=5.90 及 s_z^2=0.08911, 求出 μ_z 和 σ_a^2 的初估计.

6.4 对于英国国民生产总值按季度计量的 60 个观测值的序列 (这些数据在 [214] 中给出, 其中的一部分列在本书第五部分时间序列汇集的序列 P 中),z_t 和 ∇z_t 的自相关和偏相关函数的前 10 个值列于下表:

	z		∇z			z		∇z	
k	r_k	$\hat{\phi}_{kk}$	r_k	$\hat{\phi}_{kk}$	k	r_k	$\hat{\phi}_{kk}$	r_k	$\hat{\phi}_{kk}$
1	0.95	0.95	0.01	0.01	6	0.72	−0.01	0.00	−0.02
2	0.91	−0.01	0.06	0.06	7	0.68	−0.03	−0.26	−0.26
3	0.87	−0.03	0.13	0.13	8	0.63	−0.02	−0.19	−0.20
4	0.82	−0.02	0.03	0.02	9	0.59	0.01	0.03	0.06
5	0.77	−0.09	−0.04	−0.05	10	0.55	−0.02	−0.09	0.01

$$\bar{z}=98.5;\ s_z^2=149.6;\ \bar{w}=0.65;\ s_w^2=0.7335(w_t=\nabla z_t)$$

(a) 对于 z 和 ∇z_t 画出 ACF 和 PACF 的图形.

(b) 对于该序列识别出一个模型.

(c) 得出参数的初始估计.

6.5 用 H_t 来记美国年度生猪价格, 是由美过农业调查局 1867– 1948 共 82 个年份每年 1 月 1 日所记录的 (这些数据更详细的情况在 [162] 中给出), 定义序列 $z_t=1000\log_{10}(H_t)$, z_t 和 ∇z_t, 其前 10 个值在下表中列出:

k	$r_k(z)$	$r_k(\nabla z)$	k	$r_k(z)$	$r_k(\nabla z)$
1	0.85	0.25	6	0.46	0.14
2	0.67	−0.25	7	0.42	0.18
3	0.55	−0.35	8	0.38	0.02
4	0.51	−0.13	9	0.35	−0.07
5	0.50	0.03	10	0.32	−0.10

基于这些自相关值, 对该序列识别一个模型并求出参数的初估计.

这些数据作为序列 Q 列在本书第五部分的时间序列汇集中.

7

7.1 对于序列 z_t(如此之短并非真实的) 用 (0,1,1) 模型 $w_t = \nabla z_t = (1-\theta B)a_t$, 其中 θ=−0.5, 采用未知初始值 a_0, 计算出的值列于下表

t	z_t	$w_t=\nabla z_t$	$a_t=w_t-0.5a_{t-1}$
0	40		a_0
1	42	2	$2-0.52a_0$
2	47	5	$4+0.25a_0$
3	47	0	$-2-0.13a_0$
4	52	5	$6+0.06a_0$
5	51	−1	$-4-0.03a_0$
6	57	6	$8+0.02a_0$
7	59	2	$-2-0.01a_0$

(a) 验证表中的各项.

(b) 说明条件平方和为

$$\sum_{t=1}^{7}(a_t|-0.5, a_0=0)^2 = S_*(-0.5|0) = 144.00$$

7.2 利用习题 7.1 的数据

(a) 证明 (用最小二乘法) 使 $S_*(-0.5|a_0)$ 极小的 a_0 的值 $\hat{a}_0$ 为

$$\hat{a}_0 = \frac{(2)(0.50)+(4)(-0.25)+\cdots+(-2)(0.01)}{1^2+0.5^2+\cdots+0.01^2} = \frac{-\sum_{t=0}^{n}\theta^t a_t^0}{\sum_{t=0}^{n}\theta^{2t}}$$

这里 $a_t^0 = (a_t|\theta, a_0=0)$ 是条件值. 将 $\hat{a}_0$ 与 MA(1) 模型中精确反向预报值 $[a_0]$($[a_0]$ 的表达式由前文附录 A7.3 中方程 (A7.3.9) 给出) 相比较, 验证这两个表达式是相同的.

(b) 先将模型用反向形式写出 $w_t = (1-\theta F)e_t$, 然后递推计算 e, 证明 (a) 中得到的 a_0 值与用反向预报方法得到的 a_0 是相同的.

7.3 利用习题 7.2 中计算的 $\hat{a}_0$,

(a) 说明非条件平方和 $S(-0.5)$ 为 143.4.

(b) 对于 (0,1,1) 模型和充分大的 n, 证明

$$S(\theta) = S_*(\theta|0) - \frac{\hat{a}_0^2}{1-\theta^2}$$

7.4 对于过程 $w_t - \mu = (1-\theta B)a_t$, 证明，若序列很长，则 μ 和 θ 极大似然估计 $\hat{\mu}$ 和 $\hat{\theta}$ 的协方差矩阵近似为

$$n^{-1}\begin{bmatrix}(1-\theta)^2\sigma_a^2 & 0\\ 0 & 1-\theta^2\end{bmatrix}$$

7.5 (a) 为了得到对一个序列满意的拟合，遇到了问题，该序列的最后 16 个值记录如下

129, 135, 130, 130, 127, 126, 131, 152,
123, 124, 131, 132, 129, 127, 126, 124,

画出该序列的图，并指出困难会出现在什么地方.

(b) 用模型 $(1-\phi_1 B - \phi_2 B^2)z_t = (1-\theta B)a_t$ 拟合一组数据，收敛很慢，而且在顺次迭代过程中系数摆动很剧烈，最后得到如下标准差很大的一组估计值

$$\hat{\phi}_1{=}1.19,\ \hat{\phi}_2{=}{-0.34},\ \hat{\theta}{=}0.52$$

你能对模型的不稳定特性作出解释吗？为什么在初步识别时应该已经消除了这个问题？

(c) 拟合模型 $\nabla^2 z_t = (1-\theta_1 B-\theta_2 B^2)a_t$ 不能得到收敛结果，最终迭代得到值 $\hat{\theta}_1{=}1.81, \hat{\theta}_2{=}0.52$. 你能解释这一困难吗？

7.6 对于 ARIMA(1,1,1) 模型 $(1-\phi B)w_t = (1-\theta B)a_t$, 这里 $w_t = \nabla z_t$,

(a) 写出模型的线性化形式.

(b) 你打算如何着手进行条件非线性最小二乘计算，假设初值为 $\phi{=}0.5$, $\theta{=}0.4$, 该序列的前 9 个值列出如下：

t	z_t	t	z_t
0	149	5	150
1	145	6	147
2	152	7	142
3	144	8	146
4	150		

7.7 (a) 证明，二阶自回归模型 $\tilde{z}_t = \phi_1\tilde{z}_{t-1} + \phi_2\tilde{z}_{t-2} + a_t$ 可写成正交形式

$$\tilde{z}_t = \frac{\phi_1}{1-\phi_2}\tilde{z}_{t-1} + \phi_2\left(\tilde{z}_{t-2} - \frac{\phi_1}{1-\phi_2}\tilde{z}_{t-1}\right) + a_t$$

(这意味着，对于长序列，$\dfrac{\phi_1}{1-\phi_2}$ 的近似估计 r_1 及 ϕ_2 的近似估计 $\hat{\phi}_2=\dfrac{r_2-r_1^2}{1-r_1^2}$ 是不相关的.)

(b) 从 $\hat{\phi}_1$ 和 $\hat{\phi}_2$ 的方差 - 协方差阵或从其它条件出发，对于长序列，证明 r_1 和 $\hat{\phi}_2$ 的方差 - 协方差阵近似为

$$n^{-1}\begin{bmatrix}(1-\phi_2^2)(1-\rho_1^2) & 0\\ 0 & 1-\phi_2^2\end{bmatrix}$$

8

8.1 用试探性模型去拟合一个时间序列，下表列出所得残差序列的前 30 个：

z	残		差			
1–6	0.78	0.91	0.45	−0.78	−1.90	−2.10
7–12	−0.54	−1.05	0.68	−3.77	−1.40	−1.77
13–18	1.18	0.02	1.29	−1.30	−6.20	−1.89
19–24	0.95	1.49	1.08	0.80	2.02	1.25
25–30	0.52	2.31	1.64	0.78	1.99	1.36

画出残差图并叙述你关于模型合理性的保留意见.

8.2 用模型 $\tilde{z}_t=(1-0.6B)e_t$ 拟合 N=82 的观测值序列，得到残差序列自相关：

k	$r_k(\hat{e})$	k	$r_k(\hat{e})$
1	0.39	6	−0.13
2	0.20	7	−0.05
3	0.09	8	0.06
4	0.04	9	0.11
5	0.09	10	0.02

(a) 画出残差序列的 ACF 图形，并确定是否存在异常值.

(b) 计算 χ^2− 统计量 $\tilde{Q}$ 并检验残差序列的自相关函数，是否能从整体上说明模型的不合理性.

(c) 你将考虑用什么样的修正模型来拟合、检验？

8.3 假设 (0,1,1) 模型 $\nabla z_t=(1-\theta B)e_t$ 被用来预报一个序列，这对应于用一个指数加权滑动平均做预报，其中 θ 任意选取为 0.5. 而事实

上，这个序列可以很好地用 (0,1,2) 模型 $\nabla z_t=(1-0.9B+0.2B^2)a_t$ 来拟合。

(a) 计算由 (0,1,1) 模型所得 1 步预报误差序列的自相关函数。

(b) 说明这些自相关函数如何被用来识别序列 e_t 的模型，由此可导致对于序列 z_t 的 (0,1,2) 模型的识别.

8.4 一个有 $N=326$ 个数据的长序列被分成两半，对每一部分用一个 (1,1,0) 模型进行识别、拟合及检验. 如果这两部分的参数估计为 $\hat{\phi}^{(1)}=0.5$ 和 $\hat{\phi}^{(2)}=0.7$, 是否存在参数 ϕ 已经变化的证据？

8.5 (a) 证明由一个平稳的 AR(1) 过程 $(1-\phi B)\tilde{z}_t=a_t$ 产生的 n 个观测值的样本均值的方差为

$$\operatorname{var}[\bar{z}]\simeq\frac{\sigma_a^2}{n(1-\phi)^2}$$

(b) 一个化学过程控制在完全相同的条件下，由此得到的一批顺次产量被证明服从一个平稳的 AR(1) 过程 $(1+0.5B)\ \tilde{z}_t=a_t$ 。在给定时点进行技术改造，使得在技术改造之前第 85 个数据点的均值 $\bar{z}_1=41.0$, 残差方差 $s_{1a}=0.1012$, 而在技术改造之后第 60 个数据点的均值 $\bar{z}_2=43.5,s_{2a}^2=0.0895$ 。是否有证据表明技术改造提高了产量？

9

9.1 证明季节差分算子 $1-B^{12}$(分析月度数据是很有用的) 可以分解如下：

$$\begin{aligned}(1-B^{12})=&(1+B)(1-\sqrt{3}B+B^2)(1-B+B^2)(1+B^2)\\&\times(1+B+B^2)(1+\sqrt{3}B+B^2)(1-B)\end{aligned}$$

在单位圆上画出该表达式的零点，并且通过实际数值计算及对这些结果作图，说明按以上顺序给出的因子对应于频率 (每年的周数) 为 6,5,4,3,2,1 的正弦函数及一个常数项. (例如, 差分方程 $(1-B+B^2)x_t=0$, 对任给初值 $x_1=0,x_2=1$ 可得到 $x_3=1,x_4=0,x_5=-1$, 等等. 产生一个频率为每年 2 周的正弦波。)

9.2 用来消除月度时间序列季节性的一种方法有时常常使用，它是利用

等权的 12 个月滑动平均

$$\bar{z}_t = \frac{1}{12}(z_t + z_{t-1} + \cdots + z_{t-11})$$

(a) 利用分解 $(1-B^{12})/(1-B)=1+B+B^2+\cdots+B^{11}$, 证明 $12(\bar{z}_t - \bar{z}_{t-1}) = (1-B^{12})z_t$ 。

(b) 对于给定月份 t,z_t 对前期滑动平均的超出值为 $z_t - \bar{z}_{t-1}$ 。于是可以计算量值 u_t, 它是当前超出值与以往 k 年同期超出值平均的 u_t 比较. 证明, 可以写成

$$u_t = \left(1-\frac{B}{12}\frac{1-B^{12}}{1-B}\right)\left(1 - \frac{B^{12}}{k}\frac{1-B^{12k}}{1-B^{12}}\right)z_t$$

9.3 在 [186] 中已经说明加利福尼亚臭氧 (Azusa 由烟雾产生) 的月平均值可以用下面模型表示

$$(1-B^{12})z_t=(1+0.2B)(1-0.9B^{12})a_t, \quad \sigma_a^2=1.0$$

(a) 计算并画出 ψ 权.

(b) 计算并画出 π 权.

(c) 计算提前 3 个月和 12 个月预报的标准差.

(d) 求出最终预报函数.

9.4 从 1969 年 1 月到 1972 年 12 月 Azusa 每月臭氧平均水平如下表 (以亿分之一计):

	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1969	2.1	2.6	4.1	3.9	6.7	5.1	7.8	9.3	7.5	4.1	2.9	2.6
1970	2.0	3.2	3.7	4.5	6.1	6.5	8.7	9.1	8.1	4.9	3.6	2.0
1971	2.4	3.3	3.3	4.0	3.6	6.2	7.7	6.8	5.8	4.1	3.0	1.6
1972	1.9	3.0	4.5	4.2	4.8	5.7	7.1	4.8	4.2	2.3	2.1	1.6

利用习题 9.3 中的模型, 计算未来 24 个月的预报值 (未知的 a_t 用零近似) .

9.5 在 [185] 中 Thompson 和 Tiao 已经说明了, 在 Wisconsin 外移电话机数可由下面模型很好地描述:

$$(1-0.5B^3)(1-B^{12})z_t=(1-0.2B^9-0.3B^{12}-0.2B^{13})a_t$$

计算并画出 $w_t=(1-B^{12})z_t$ 的自相关函数，滞后为 1,2,···,24.

9.6 一个 N=41 的银行存款的季度数据序列，它的各种差分序列的前 12 个自相关在下表列出：

	k													
	1	2	3	4	5	6	7	8	9	10	11	12	$\bar{w}$	s_w^2
$r_k(z)$	0.88	0.83	0.72	0.67	0.56	0.53	0.42	0.41	0.32	0.29	0.21	0.18	152.0	876.16
$r_k(\nabla z)$	−0.83	0.68	−0.70	0.70	−0.65	0.62	−0.70	0.69	−0.60	0.55	−0.60	0.62	2.46	97.12
$r_k(\nabla_4 z)$	0.21	0.06	0.15	−0.34	−0.10	0.03	−0.18	−0.13	0.02	−0.04	0.10	0.22	10.38	25.76
$r_k(\nabla\nabla_4 z)$	−0.33	−0.06	0.36	−0.45	0.08	0.22	−0.10	−0.05	0.14	−0.12	−0.05	0.28	0.42	38.19

画出自相关函数图，对该序列进行模型识别，并计算参数和 σ_a^2 的初步估计.

9.7 在分析季节性产品销售量的对数时间序列时，发现由于月度和季节性变化，需要作形如 $w_t=\nabla\nabla_{12}\ln(z_t)$ 的差分，以使序列平稳. 下表是包含 n=102 个月度观测值序列 w_t 的前 48 个自相关函数：

滞后						自	相	关				
1–12	−0.39	−0.24	0.17	0.21	−0.27	−0.03	0.26	−0.10	−0.20	0.07	0.44	−0.58
13–24	0.09	0.17	0.01	−0.24	0.16	0.04	−0.12	−0.01	0.11	0.08	−0.33	0.28
25–36	0.01	−0.14	−0.02	0.18	−0.13	0.04	−0.01	0.10	−0.13	−0.09	0.27	−0.22
37-48	0.00	0.09	0.02	−0.18	0.17	−0.05	0.0	−0.06	0.06	0.06	−0.13	0.11

$$\overline{w}=0.241 \qquad s_w^2=106.38$$

对这个序列识别一个 (或多个) 适当的模型，并求出参数的初步估计.

10

10.1 设 X_t 是一个煤气炉的甲烷馈入速度 (单位：立方英尺 / 秒)，Y_t 是输出气体中二氧化碳的百分含量. 下列传递函数模型

$$(1)\ Y_t=10+\frac{25}{1-0.7B}X_{t-1}$$

$$(2)\ Y_t=10+\frac{22-12.5B}{1-0.85B}X_{t-2}$$

$$(3)\ Y_t=10+\frac{20-8.5B}{1-1.2B+0.4B^2}X_{t-3}$$

(a) 证明上述模型是稳定的.

(b) 计算稳态增益 g(用适当的单位表示) .

10.2 对于习题 10.1 中的各个模型，用差分方程计算并画出对应下列输入的响应

(a) 在时间 $t=0$ 使用单位脉冲输入 (0,1,0,0,0,0,$\cdots$)。

(b) 在时间 $t=0$ 使用阶跃输入 (0,1,1,1,1,$\cdots$)。

(c) 在时间 $t=0$ 使用斜面输入 (0,1,2,3,4,5,$\cdots$)。

(d) 在时间 $t=0$ 使用周期输入 (0,1,0,-1,0,1,$\cdots$)。

估计模型 (3) 对于阶跃响应的周期和衰减因子.

10.3 利用方程 (10.2.8) 计算出习题 10.1 中每个模型的脉冲权 v_j, 并检验它们与习题 10.2(a) 所求得的脉冲响应是相同的.

10.4 把习题 10.1 中的模型用 ∇ 形式表示.

10.5 (a) 计算并画出两输入系统

$$Y_t = 10 + \frac{6}{1-0.7B}X_{1,t-1} + \frac{8}{1-0.5B}X_{2,t-2}$$

的响应. 输入序列是正交和随机化的，列出如下:

t	X_{1t}	X_{2t}	t	X_{1t}	
0	0	0	5	1	-1
1	-1	1	6	1	1
2	1	-1	7	-1	-1
3	-1	-1	8	-1	1
3	1	1			

(b) 分别计算 Y 对于 X_1 和 X_2 的增益 g_1 和 g_2, 并将模型用 ∇ 形式表示.

1 1

11.1 如果两个序列可以用 ψ 权的形式表示为

$$y_t = \psi_y(B)a_t \qquad x_t = \psi_x(B)a_t$$

(a) 证明互协方差的生成函数

$$\gamma^{xy}(B) = \sum_{k=-\infty}^{\infty} \gamma_{xy} B^k$$

是由 $\psi_y(B)\psi_x(F)\sigma_a^2$ 给出的.

(b) 如果

$$y_t = (1-\theta B)a_t\ ,\qquad x_t = (1-\theta'_1 B-\theta'_2 B^2)a_t$$

利用前面的结果得出 y_t 和 x_t 之间的互协方差函数.

11.2 对输入序列 x_t, 在估计其预白噪化变换 $\theta_x^{-1}(B)\phi_x(B)x_t=\alpha_t$ 后, 计算经变换后的输出 $\beta_t=\theta_x^{-1}(B)\phi_x(B)y_t$, 并得到互相关系数 $\gamma_{\alpha\beta}(k)$ 如下:

k	$r_{\alpha\beta}(k)$	k	$r_{\alpha\beta}(k)$
0	0.05	5	0.24
1	0.31	6	0.07
2	0.52	7	−0.03
3	0.43	8	0.10
4	0.29	9	0.07

其中 $\hat{\sigma}_\alpha=1.26, \hat{\sigma}_\beta=2.73, n=187$.

(a) 求出互相关标准差的近似值.

(b) 计算脉冲响应权 v_j 的粗估计.

(c) 给出传递函数的一个模型及其参数的粗估计.

11.3 经常遇到这样的情形: 使用者估计出传递函数模型 $y_t=\omega(B)\delta^{-1}(B)B^b x_t$, 并想确认稳态增益是否有意义.

(a) 对于一阶系统

$$y_t=\frac{\omega_0}{1-\delta B}x_{t-1}$$

证明估计值 $\hat{g}=\hat{\omega}_0/(1-\hat{\delta})$ 的近似标准差由下式给出

$$\frac{\hat{\sigma}^2(\hat{g})}{\hat{g}^2}\simeq\frac{\text{var}[\hat{\omega}_0]}{\hat{\omega}_0^2}+\frac{\text{var}[\hat{\delta}]}{(1-\hat{\delta})^2}+\frac{2\text{cov}[\hat{\omega}_0,\hat{\delta}]}{\hat{\omega}_0(1-\hat{\delta})}$$

(b) 若 $\hat{\omega}_0=5.2, \hat{\delta}=0.65$, $\hat{\sigma}(\hat{\omega}_0)=0.5$, $\hat{\sigma}(\hat{\delta})=0.1, \text{cov}[\hat{\omega}_0,\hat{\delta}]=0.025$, 计算 $\hat{g}$ 和 $\hat{\sigma}(\hat{g})$ 的近似值.

11.4 考虑模型

$$Y_t=\beta_1X_{1,t}+\beta_2X_{2,t}+N_t$$

这里 N_t 是非平稳误差项, 并且 $\nabla N_t=a_t-\theta z_{t-1}$. 证明模型可

以写成形如

$$Y_t-\bar{Y}_{t-1}=\beta_1(X_{1,t}-\bar{X}_{1,t-1})+\beta_2(X_{t,t}-\bar{X}_{2,t-1})+a_t$$

其中 $\bar{Y}_{t-1},\bar{X}_{1,t-1},\bar{X}_{2,t-1}$ 是指数加权滑动平均．例如

$$\bar{Y}_{t-1}=(1-\theta)(Y_{t-1}+\theta Y_{t-2}+\theta^2 Y_{t-3}+\cdots)$$

可以看到，用极大似然方法拟合该非平稳模型，等价于用最小二乘法去拟合由局部修正指数加权滑动平均所得独立和非独立变量之间的离差，可参见第 9.4.2 节的相应思想．

11.5 分别考虑习题 6.3 和 6.4 中的失业和国民生产总值数据，即第五部分时间序列汇集的序列 P．建立 (识别、估计并检验) 一个传递函数模型，并用 GDP 序列 X_t 作为输入以辅助解释对数失业人数序列 Y_t 的变化．

12

12.1 在烟雾产生臭氧的月度数据 y_t 的分析 [59] 中，考虑如下两种可能的"干预"I_1 和 I_2 如下：

I_1: 在 1960 年初，由于金州高速公路开通造成交通状况改变，以及要求减少地方销售汽油中碳氢化合物含量的有关法律生效．

I_2: 1966 年要求所有新汽车安装新发动机设备的法律生效．在这种干预情形，考虑到众所周知的事实在夏天和冬天的月份里，烟雾现象是不同的．在数据的试探性分析中，用到下列模型

$$y_t=\omega_1\xi_{1t}+\frac{\omega_2}{1-B^{12}}\xi_{2t}+\frac{\omega_2}{1-B^{12}}\xi_{3t}+\frac{1(1-\theta B)(1-\Theta B^{12})}{1-B^{12}}a_t$$

其中

$$\xi_{1t}=\begin{cases}0, & t<1960\text{ 年 1 月}\\ 1, & t>1960\text{ 年 1 月}\end{cases}$$

$$\underset{(\text{夏季月份})}{\xi_{2t}=\begin{cases}0, & t<1966\text{ 年 1 月}\\ 1, & t>1966\text{ 年 1 月}\end{cases}},\quad \underset{(\text{冬季月份})}{\xi_{3t}=\begin{cases}0, & t<1966\text{ 年 1 月}\\ 1, & t>1966\text{ 年 1 月}\end{cases}}$$

(a) 说明该模型考虑到

(1) 在 1960 年 1 月可能由 I_1 产生大小为 ω_1 的阶跃变化.

(2) 年跃度为 ω_2 的阶跃函数，这是考虑到随着装有新发动机设备汽车的逐渐涌入而产生的夏季影响.

(3) 年跃度为 ω_3 的阶跃函数，这是考虑到随着装有新发动机设备汽车的逐渐涌入而产生的冬季影响.

(b) 叙述你将采取什么步骤来检验模型表示的合理性.

(c) 假设在 (b) 之后你已经满意了，由下面的结果你将得出什么结论？（下面给了估计值，括号里面是它们对应的标准差.）

$$\begin{array}{lllll} \hat{\omega}_1 = -1.09, & \hat{\omega}_2 = -0.25, & \hat{\omega}_3 = -0.07 & \hat{\theta} = -0.24, & \hat{\Theta} = 0.55 \\ (\pm 0.13) & (\pm 0.07) & (\pm 0.06) & (\pm 0.03) & (\pm 0.04) \end{array}$$

12.2 一个一般的传递函数模型

$$y_t = \sum_{j=1}^{k} \omega_j(B)\delta_j^{-1}(B)\xi_{jt} + \theta(B)\phi^{-1}(B)a_t$$

它可以包含本身为时间序列的输入变量 ξ_j 和作为指示变量的其它输入变量 ξ_i, 后者可以估计 (并消除) 习题 12.1 中所述一类干预产生的影响，这在销售数据的分析中经常用到. 令 $\xi_t^{(T)}$ 为一指示变量，它在时间 T 采取单位脉冲的形式，即

$$\xi_t^{(T)} = \begin{cases} 0 & t \neq T \\ 1 & t = T \end{cases}$$

为了说明起见，考虑模型

(1) $y_t = \dfrac{\omega_1 B}{1 - \delta B}\xi_t^{(T)}$ (其中$\omega_1 = 1.00, \delta = 0.50$)

(2) $y_t = \left(\dfrac{\omega_1 B}{1 - \delta B} + \dfrac{\omega_2 B}{1 - B}\right)\xi_t^{(T)}$

(其中$\omega_1 = 1.00, \delta = 0.50, \omega_2 = 0.30$)

(3) $y_t = \left(\omega_0 + \dfrac{\omega_1 B}{1 - \delta B} + \dfrac{\omega_2 B}{1 - B}\right)\xi_t^{(T)}$

(其中$\omega_0 = 1.50, \omega_1 = -1.00, \delta = 0.50, \omega_2 = -0.50$)

递推计算各个模型在时间 $t = T, T+1, T+2, \cdots$ 的响应 y_t, 并评述它们在估计和 / 或消除由于例如广告运动，促销以及价格变化等现象影响下可能产生的作用.

12.3 对于习题 6.5 的生猪价格序列，经过探索性识别得到一个模型，用最小二乘法拟合该模型求出最终参数估计，并且对残差进行模型诊断检验. 特别地，对于 1932- 1935 和 1942-1948 的这些数据，考虑外部因素或者干预事件. 如果适合的话，设定一个有指示输入变量的传递函数模型来改进基本的 ARIMA 模型，并且估计这个设定的模型

1 3

13.1 在一个化学过程中，把控制变量 (煤气速度)X_t 固定在它的标准参考原点，在此期间顺次记录了粘度 N_t 的 30 个数据，如下表所示

时刻	粘	度								
1–10	92	92	96	96	96	98	98	100	100	94
11–20	98	88	88	88	96	96	92	92	90	90
21–30	90	94	90	90	94	94	96	96	96	96

重新构造并画出误差序列 (对于目标值的偏离)ϵ_t 以及调整值序列 x_t. 假设在此期间最优反馈控制方案

$$x_t = -10\epsilon_t + 5\epsilon_{t-1} \tag{1}$$

被应用于给定动态模型

$$\tilde{y}_t = 0.5\tilde{y}_{t-1} + 0.10x_{t-1} \tag{2}$$

以及由

$$\epsilon_t = \epsilon_{t-1} + \nabla N_t + \tilde{y}_t \tag{3}$$

得到的误差信号. 你的计算顺序应该按照 (2),(3) 和 (1), 并且在开始时假设 $\epsilon_1 = 0, \tilde{y}_t = 0, x_1 = 0$. 你能够设计一个由 N_t 计算 ϵ_t 更直接的方法吗？

13.2 已知下列扰动和传递函数模型的组合:

$$(1)\qquad \nabla N_t = (1-0.7B)a_t$$
$$(1-0.4B)\mathcal{Y}_t = 5.0X_{t-1+}$$
$$(2)\qquad \nabla N_t = (1-0.5B)a_t$$
$$(1-1.2B+0.4B^2)\mathcal{Y}_t = (20-8.5B)X_{t-1+}$$
$$(3)\qquad \nabla^2 N_t = (1-0.9B+0.5B^2)a_t$$
$$(1-0.7B)\mathcal{Y}_t = 3X_{t-1+}$$
$$(4)\qquad \nabla N_t = (1-0.7B)a_t$$
$$(1-0.4B)\mathcal{Y}_t = 5.0X_{t-2+}$$

(a) 由扰动和传递函数模型的每组合，设计最小均方误差反馈控制方案。

(b) 对于情况 (4)，推导误差 ϵ_t 的表示式，并且导出用 σ_a^2 表示的方差。

(c) 对于情况 (4)，设计一个适合于操作者做出手工控制行动的诺曼图.

13.3 在一个工业废料处理工厂，废料流入的强度 z_t 满足关系式

$$\tilde{Y}_t = \frac{0.3B}{1-0.2B}\tilde{z}_t$$

废料流入强度的增加可以由工厂增大使用一种化学物质的流量 u_t 进行补偿， u_t 对 Y_t 的影响可由模型

$$\tilde{Y}_t = \frac{21.6B^2}{1-0.7B}\tilde{u}_t$$

表示。证明，利用控制方程

$$\tilde{u}_t = -\frac{0.3}{21.6}\left[\frac{(0.7-0.2B)(1-0.7B)}{(1-0.2B)(1-0.5B)}\right]\tilde{z}_t$$

即由 $u_t = 0.7\tilde{u}_{t-1} - 0.1\tilde{u}_{t-2} - 0.0139(0.7\tilde{z}_t - 0.69\tilde{z}_{t-1} + 0.14\tilde{z}_{t-2})$ 可得到最小均方误差反馈控制.

13.4 在下面扰动和传递函数模型

$$\nabla N_t = a_t$$
$$(1-\delta B)\mathcal{Y}_t = \omega_0 X_{t-1+} - \omega_1 X_{t-2+}$$

的基础上，实施一个试验性反馈控制方案，得到调整序列 x_t 和误差序列 ϵ_t 。据信误差模型是比较准确的，但是传递函数模型参数的准确性令人怀疑。

(a) 下表给出序列 x_t 和 ϵ_t 的前 10 个值

t	x_t	ϵ_t	t	x_t	ϵ_t
1	25	−7	6	−30	1
2	42	−7	7	−25	3
3	3	−6	8	−25	4
4	20	−7	9	20	0
5	5	−4	10	40	−3

对 δ=0.5,ω_0=0.3,ω_1=0.2 以及任意初值 y_t^0 和 x_0^0 计算残差 a_t(t=2, 3,⋯,10) 。

(b) 计算 y_1^0 和 x_0^0 的值 $y_1, \hat{x}_0$, 以使平方和

$$\sum_{t=2}^{10}(a_t|\delta=0.5,\omega_0=0.3,\omega_1=0.2,y_1^0,x_0^0)^2$$

最小，并计算该最小平方和。

13.5 考虑 [55] 中的一个系统，过程的传递函数为 gB, 噪声模型是 $(1-B)N_t=(1-\theta B)a_t$, 使得输出误差满足

$$(1+B)\epsilon_t = g(1-B)X_{t-1+}+(1-\theta B)a_t$$

假设系统被如下已知的离散“积分”控制器所控制:

$$(1-B)X_{t+} = -c\epsilon_t$$

(a) 证明，输出误差服从 ARIMA(1,1) 过程

$$(1-\phi B)\epsilon_t = (1-\theta B)a_t \quad , \qquad \phi = 1-gc$$

因此，利用数据从一个试验性方案估计 g 和 θ 的问题，等价于估计该模型的参数。

(b) 还需证明，最优方案使得 $c = c_0 = (1-\theta)g$, 而且，如果在收集数据时用到的试验性方案恰好是最优的，那么，$1-\theta$ 和 g 就不能分别估计。

参考文献

[1] Abraham,B., "Missing observations in time series", *Comm. Statist.*,**A10**,1643-1653,1981.

[2] Abraham, B., and G. E. P. Box, "Deterministic and forecast -adaptive time-dependent models, "*Appl. Statist.* , **27** , 120-130,1978.

[3] Akaike,H.,"A new look at the statistical model identification", *IEEE Trans. Automatic Control,***AC-19**,716-723,1974.

[4] Akaike,H.,"Canonical correlation analysis of time series and the use of an information criterion",*in Systems Identification: advances and Case Studies,* ed. R.K. Mehra and D.G. Lainiotis,Academic Press,New York,1976,pp.27-96.

[5] Anderson,B. D. O. ,and J. B. Moore, *Optimal Filtering,* Prentice Hall,Englewood Cliffs,N.J.,1979.

[6] Anderson, R. L. , "Distribution of the serial correlation coefficient", *Ann. Math. Statist.* , **13**, 1-13, 1942.

[7] Anderson, T. W. ,*An Introduction to Multivariate Statistical Analysis,* 2nd ed.,Wiley,New York,1984.

[8] Anderson. T. W. , *The Statistical Analysis of Time Series,* Wiley, New York, 1971.

[9] Anscombe,F.J."Examination of residuals",*Proc. 4th Berkeley Symp.*,**1**,1-36,1961.

[10] Anscombe,F.J.,and J.W.Tukey,"The examination and analysis of residuals",*Technometrics,***5**,141-160,1963.

[11] Ansley,C.F.,"An algorithm for the exact likelihood of a mixed autoregressive moving average process",*Biometrika,***66**,59-65, 1979.

[12] Ansley ,C. F. , and R. Kohn. "Exact likelihood of vector autoreg -ressive-moving average process with missing or aggregated data",*Biometrika,***70**,275-278,1983.

[13] Ansley,C.F.,and R.Kohn,"Estimation ,filtering,and smoothing in state space models with incompletely specified initial conditions", *Ann.Statist.*,**13**, 1286-1316,1985.

[14] Ansley, C. F. , and P. Newbold, "Finite sample properties of estimators for autoregressive moving average models" , *J. Econometrics* , **13**, 159–183, 1980.
[15] Aström,K. J. ,and T. Bohlin,"Numerical identification of linear dynamic systems from normal operating records",in *Threory of Self-Adaptive Control Systems.* ed. P. H. Hammond, Plenum Press,New York,1966,pp.96–111.
[16] Aström,K.J.,and B.Wittenmark,*Computer Controlled Systems,* Prentice Hall.Englewood Cliffs,N.J.,1984.
[17] Bachelier.L.,"Théorie de la spéculation",*Ann.Sci. École Norm. Sup.*,Ser.3,**17**,21–86,1900.
[18] Bacon, D.W., "Scasonal time series", Ph.D.thesis,University of Wisconsin-Madison,1965.
[19] Bagshaw,M.,and R.A.Johnson,"Sequential Procedures for detecting parameter changes in a time-series mode." ,*J. Amer, Statist.Assoc.*,**72**,593–597,1977.
[20] Baillie,R.T.,"The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors",*J.Amer, Statist. Assoc.*,**74**,175–184,1979.
[21] Barnard,G. A., "Control charts and stochastic processes," *J. Roy. Statist.Soc.*,**B21**,239–257,1959.
[22] Barnard,G.A.,"Statistical inference," *J.Roy. Statist. Soc.* ,**B11**, 115–139,1949.
[23] Barnard,G.A.,"The logic of least squares",*J.Roy. Statist. Soc.,* **B25**,124–127,1963.
[24] Barnard,G. A., G. M. Jenkins, and C.B.Winsten, "Likelihood inference and time series",*J.Roy. Statist. Soc.*,**A125**,321–352, 1962.
[25] Bartlett,M.S.,"On the theoretical specification and sampling properties of autocorrelated time-series", *J.Roy.Statist. Soc.* , **B8**,27–41,1946.
[26] Bartlett.M.S.,*Stochastic Processes,*Cambridge University Press, Cambridge,1955.
[27] Bell,W.,"A note on overdifferencing and the equivalence of seasonal time series models with monthly means and models with $(0,1,1)_{12}$ seasonal parts when $\Theta=1$,*J.Business Econom. Statist.,* **5**,383–387,1987.

[28] Bell,W.R.,and S.C. Hilmer,"Modeling time series with calendar variation",*J. Amer. Statist. Assoc.*,**78**,526–534,1983.

[29] Bell,W.R.,and S.Hilmer,"Initalizing the Kalman filter in the nonstationary case",*ASA Proc.Business and Economic Statistics Section,* 1987,pp.693–697.

[30] Bergh, L. G. , and J.F.MacGregor, "Constrained minimum variance controllers:internal model structure and robustness properties", *Ind. Eng. Chem. Res.,***26**,1558–1564,1987.

[31] Bhattacharyya ,M. N. , and A. P. Layton, "Effectiveness of seat belt legislation on the Queensland road toll—an Australian case study in intervention analysis", *J. Amer. Statist. Assoc.,* **74** , 596–603,1979.

[32] Birnbaum,A.,"On the foundations of statistical inference", *J. Amer. Statist. Assoc.,***57**,269–306,1962.

[33] Booth, G. W. , and T. I. Peterson, "Non-linear estimation", *IBM Share Program Pa .No.687 WL NLI,*1958.

[34] Box, G. E. P. , "Use and abuse of regression" , *Technometrics,* **8**, 625–629,1966.

[35] Box,G.E. P.,"Sampling and Bayes' inference in scientific modeling and robustness", *J.Roy. Statist. Soc.,***A143** 383–430,1980.

[36] Box,G.E. P., "Feedback control by manual adjustment", *Quality Engrg.*,**4**(1),143–151,1991.

[37] Box , G. E. P. , "Bounded adjustment charts" , *Quality Engrg.* , **4**(a),331–338,1991.

[38] Box, G. E. P.,and D. R. Cox,"An analysis of transformations", *J. Roy. , Statist. Soc.*, **B26**, 211–243, 1964.

[39] Box, G. E. P.,aand N. R. Draper,"The Bayesian estimation of common parameters from several responses",*Biometrika,***52**, 355–365,1965.

[40] Box, G. E. P., and W. G. Hunter,"The experimental study of physical mechamisms",*Technometrics,***7**,23–42,1965.

[41] Box, G. E. P., W. G. Hunter,and J. S. Hunter, *Statistics for Experimenters,* Wiley, New York,1978.

[42] Box, G. E. P.,and G. M. Jenkins,"Further contributions to adaptive quality control :simultaneous estimation of dynamics: non-zero costs", *Bull. Internat. Statist. Inst. , 34th Session,* Ottawa, Canada, 1963., pp. 943–974.

[43] Box,.G. E. P.,and G. M. Jenkins, "Discrete models for feedback and feed-forward control", in *The future of Statistics,* ed. D. G. Watts, Academic Press, New York,1968,pp.201–240.

[44] Box,G. E. P., and G. M. Jenkins,"Some recent advances in forecasting and control I",*Appl. Statist.*,**17**,91–109,1968.

[45] Box, G. E. P., and G. M. Jenkins, "Discrete models for forecasting and control", in *Encyclopaedia of Linguistics, Information and Control,* ed. A. R. Meaham and R. A. Hudson, Pergamon Press, Elmsford, N.Y.,1969,p.162.

[46] Box, G. E. P., and G. M. Jenkins,"Some statistical aspects of adaptive optimization and control",*J. Roy. Statist. Soc.*, **B24**, 297–331,1962.

[47] Box, G. E. P,and G. M. Jenkins,"Mathematical Models for adaptive control and optimization", *AIChE J.Chem. E Symp. Ser.*,**4**, 61,1965.

[48] Box, G. E. P.,and G. M. Jenkins,*Time Series Analysis: Forecasting and Control,*revised edition, Holden-Day, San Francisco, 1976.

[49] Box,G. E. P., G. M. Jenkins, and D. W. Bacon,"Models for forecasting seasonal and nonseasonal time series", in *Spectral Analysis of Time Series,* ed. B. Harris,Wiley,New York,1967,pp.271 –311.

[50] Box,G. E. P., G. M. Jenkins, and J. F. MacGregor,"Some recent advances in forecasting and control,part II",*Appl. Statist.*, **23**,158–179,1974.

[51] Box,G. E. P., G. M. Jenkins,and D. W. Wichern,"Least squares analysis with a dynamic model" ,*Technical Report 105,* Department of Statistics,University of Wisconsin –Madison,1967.

[52] Box,G. E. P.,and T. Kramer,"Statistical process monitoring and feedback adjustments — a discussion", *Technometrics,* **34**, 251–285,1992.

[53] Box,G. E. P.,and A. Lucéno,"Charts for optimal feedback control with recursive sampling and adjustment",*Report 89,*Center for Quality and Productivity Improvement,University of Wisconsin–Madison,1993.

[54] Box,G. E. P.,and J. F. MacGregor,"The analysis of closed-loop dynamic stochastic systems",*Technometrics,***16**,391–398, 1974.

[55] Box,G. E. P.,and J. F. MacGregor,"Parmeter estimation with colsed-loop operating data",*Technometrics*,**18**,371–380,1976.

[56] Box,G. E. P.,and D. A. Pierce,"Distribution of residual autocorrelations in autoregressive-integrated moving average time series models",*J Amer. Statist. Assoc.*,**65**,1509–1526,1970.

[57] Box,G. E. P.,and J. Ramírez,"Cumulative score charts", *Quality Reliability Engrg.*,**8**,17–27,1992.

[58] Box,G. E. P.,and G. C. Tiao, *Bayesian Inference* ,Addison-Wesley,Reading. Mass.,1973.

[59] Box,G. E. P,and G. C. Tiao,"Intervention analysis with application to economic and environmental problems",*J. Amer. Statist. Assoc.***70**,70–79,1975.

[60] Box,G. E. P.,and G. C. Tiao,"Comparison of forecast and actuality",*Appl. Statist.*,**25**,195–200,1976.

[61] Box,G. E. P.,and G. C. Tiao,"Multiparameter problems from a Bayesian point of view",*Ann. Math. Statist.*,**36**,1468–1482, 1965.

[62] Briggs, P. A. N.,P. H. Hammond, M. T. G. Hughes,and G. O. Plumb,"Correlation analysis of process dynamics using pseudo-random binary test perturbations",*Inst. Mech. Eng., Advances in Automatic Control.*, Paper 7,Nottingham. U.K.,April 1965.

[63] Brockwell,P. J. , and R. A. Davis , *Time Series : Theory and Methods*,Springer-Verlag,New York,1987.

[64] Brown, R. G.,*Smoothing,Forecasting and Prediction of Discrete Time Series,* Prentice Hall,Englewood Cliffs, N.J.,1962.

[65] Brown. R. G.,and R. F. Meyer,"The fundamental theorem of exponential smoothing",*Operations Res.*,**9**,673–685,1961.

[66] Brubacher,S.R.,and G. Tunnicliffe Wilson,"Interpolating time series with applications to the estimation of holiday effects on electricity demand",*Appl. Statist.*,**25**,107–116,1976.

[67] Bruce. A.G.,and R. D. Martin,"Leave-k-out diagnostics for time series"(with discussion),*J. Roy. Statist. Soc.*,**B51**, 363–424,1989.

[68] Chan,N.H.,and C.Z.Wei,"Limiting distributions of least squares estimates of unstable autoregressive processes",*Ann. Statist.*, **16**,367–401,1988.

[69] Chang, I., G.C. Tiao, and C. Chen, "Estimation of time series parameters in the presence of outliers", *Technometrics*, **30** ,

193–204, 1988.
[70] Chatfield, C., "Inverse autocorrelations", *J. Roy. Statist. Soc.*, **A142**.363–377,1979.
[71] Cleveland, W.S., "The inverse autocorrelations of a time series and their applications", *Technometrics,* **14**,277-293,1972.
[72] Cooper,D. M.,and E. F. Wood, "Identifying multivariate time series models", *J. Time Ser. Anal.,* **3**,153–164,1982.
[73] Cornish,E.A., "The multivariate t-distribution associated with a set of normal sample deviates", *Austral. J. Phys.,* **7**, 531,1954.
[74] Cox,D.R., "Prediction by exponentially weighted moving averages and related methods", *J. Roy. Statist. Soc.*,**B23**,414–422,1961.
[75] Damsteth, E., "Interpolating missing values in a time series", *Scand. J. Statist.,* **7**,33–39,1980.
[76] Daniel, C., "Use of half-normal plots in interpreting factorial two-level experiments", *Technometrics,* **1**,311–341,1959.
[77] Daniels, H.E., "The approximate distribution of serial correlation coefficients", *Biometrika,* **43**,169–185,1956.
[78] Davies,N.,C.M.Triggs,and P.Newbold, "Significance levels of the Box-Pierce portmanteau statistic in finite samples", *Biometrika,* **64**,517–522,1977.
[79] Deming,W.E., *out of the Crisis,* Center for Advanced Engineering Study,MIT, Cambridge,Mass.,1986.
[80] Dent,W.,and A.S.Min, "A Monte Carlo study of autoregressive integrated moving average processes", *J. Econometrics,* **7**,23–55, 1978.
[81] Dickey,D.A.,W.R.Bell,and R.B.Miller, "Unit roots in time series models:tests and implications", *Amer. Statist.,* **40**,12–26,1986.
[82] Dickey,D.A.,and W.A.Fuller, "Distribution of the estimates for autoregressive time series with a unit root", *J.Amer. Statist. Assoc.* ,**74**,427–431,1979.
[83] Dickey,D.A.,and S. G. Pantula, "Determining the order of differencing in autoregressive processes", *J. Business Econom. Statist.,* **5**,455–461,1987.
[84] Doob,J.L., *Stochastic Processes,* Wiley,New York,1953.
[85] Dudding,B.P.,and W.J.Jennet, "Quality control charts", *British Standard 600R,* 1942.
[86] Dunnett, C. W., and M. Sobel, "A bivariate generalization of

Student's t-distribution,with tables for certain special cases", *Biometrika*,**41**,153–169,1954.

[87] Durbin, J., "The fitting of time-series models", *Rev. Internat. Statist. Inst.*,**28**,233–244,1960.

[88] Durbin,J.,"Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables", *Econometrica,***38**,410–421,1970.

[89] Dubin,J.,"An alternative to the bounds test for testing for serial correlation in least-squares regression",*Econometrica,***38**,422 – 429,1970.

[90] Fearn ,T., and P.I. Maris , "An application of Box - Jenkins methodology to the control of gluten addition in a flour mill" , *Appl. Statist.*,**40**,477–484,1991.

[91] Fisher,R.A.,*Statistical Methods and Scientific Inference,* Oliver & Boyd, Edinburgh,1956.

[92] Fox, A. J., "Outliers in time series", *J. Roy. Statist. Soc.*, **B34**, 350–363,1972.

[93] Fuller,W.A.,*Introduction to Statistical Time Series,* Wiley,New York,1976.

[94] Gardner, G. , A. C. Harver, and G. D. A. Phillips,"Algorithm AS 154: an algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman fitering", *Appl. Statist.* , **29**, 311–322, 1980.

[95] Gersch,W.,and G.Kitagawa,"The prediction of time series with trends and seasonalities",*J. Business Econom. Statist.*,**1**, 253–264,1983.

[96] Godfrey,L.G.,"Testing the adequacy of a time series model", *Biometrika*,**66**,67–72,1979.

[97] Gray.H.L.,G.D. Kelley,and D.D. McIntire,"A new approach to ARMA modelling",*Comm. Statist.*,**B7**,1–77,1978.

[98] Grenander, U., and M. Rosenblatt, *Statistical Analysis of Sta -tionary Time Series,*Wiley,New York,1957.

[99] Hald, A., *Statistical Theory with Engineering Applications*, Wiley,New York,1952.

[100] Hannan,E.J.,*Time Series Analysis,*Methuen,London,1960.

[101] Hannan,E.J.,and M.Deistler,*The Statistical Theory of Linear Systems,*Wiley,New York,1988.

[102] Hannan, E. J. ,W. M. Dunsmuir,and M. Deistler, "Estimation of

vector ARMAX models", *J. Multivariate Anal.*, **10**, 275-295, 1979.

[103] Hannan, E. J., and J.Rissanen, "Recursive estimation of mixed autoregressive moving avernage order", *Biometrika,* **69,** 81-94, 1982.

[104] Harris, T. J., J. F.MacGregor, and J.D.Wright, "An overview of discrete stochastic controllers:generalized PID algorithms with dead-time compensation", *Canad. J. Chem. Engrg.*, **60**, 425–432, 1982.

[105] Harrison, P. J., "Short-term sales forecasting", *Appl. Statist.*, **14**, 102–139, 1965.

[106] Harvey, A.C., "Finite sample prediction and overdifferencing", *J. Time Ser. Anal.*, **2**, 221–232, 1981.

[107] Harvey, A.C., and G.D.A. Phillips, "Maximum likelihood estimation of regression models with autoregressive-moving avernage disturbances", *Biometrika,* **66**, 49–58, 1979.

[108] Harvey, A.C., and R.G.Pierse, "Estimating missing observations in economic time series", *J. Amer. Statist. Assoc.*, **79**, 125–131, 1984.

[109] Havey, A.C., and P.H.J.Todd, "Forecasting economic time series with structural and Box-Jenkins models:a case study", *J. Business Econom. Statist.*, **1**, 299–307, 1983.

[110] Hillmer, S.C., and G.C. Tiao, "Likelihhod function of stationary multiple autoregressive moving average models", *J. Amer. Statist. Assoc.*, **74**, 652–660, 1979.

[111] Hillmer, S.C., and G.C.Tiao, "An ARIMA-model-based approach to seasonal adjustment", *J. Amer. Statist. Assoc.*, **77**, 63-70, 1982.

[112] Holt, C.C., "Forecasting trends and seasonals by exponentially weighted moving averages", *O.N.R. Memorandum 52,* Carnegie Institute of Technology, Pittsburgh, Pa., 1975.

[113] Holt, C. C., F.Modigliani, J. F. Muth, and H.A.Simon, *Planning Production, Inventories and Work Force,* Prentice Hall. Englewood Cliffs. N.J., 1963.

[114] Hougen, J.O., *Experience and Experiments with Process Dynamics,* Chmical Engineering Progress Monograph Series, **60**(4), 1964.

[115] Hutchinson,A.W.,and R.J.Shelton,"Measurement of Dynamic characteristics of full-scale plant using random perturbing signals:an application to a refinery distillation column" , *Trans. Inst. Chem. Eng.*,**45**,334–342,1967.
[116] Ishikawa,K.,*Guide to Quality Control,*Asian Productivity Organization,Tokyo,1976.
[117] Jeffreys,H.,*Theory of Probability,3rd.,*Clarendon Press, Oxford 1961.
[118] Jenkins,G.M.,"Tests of hypothese in the linear autoregressive model,I",*Biometrika,***41**,405–419,1954;II,Biometrika,**43**,186-199, 1956.
[119] Jenkins,G.M.,contribution to the discussion of the paper "Relationships between Bayesian and confidence limits for predictors", by A.R. Thatcher,*J. Roy. Statist. Soc.,***B26**,176–210,1964.
[120] Jenkins,G.M.,"The interaction between the muskrat and mink cycles in North Canada",*Proc. 8th International Biometric Conference*, Editura Academiei Republicii Socialiste Romania, Bucharest,1975,pp.55–71.
[121] Jenkins,G.M.,*Practical Experiences with Modelign and Forecasting Time Series,*Gwilym Jenkins & Partners Ltd.,Jersey, Channel Islands,1979.
[122] Jenkins,G.M.,and D.G. Watts,*Spectral Analysis and Its Applica tions,*Holden–Day,San Francisco,1968.
[123] Jones,R.H.,"Maximum likelihood fitting of ARMA models to time series with missing observations",*Technometrics,***22**,389–395, 1980.
[124] Kalman,R.E.,"A new approach to linear filtering and prediction problems",*J. Basic Engrg.,***82**,35–45,1960.
[125] Kalman,R.E.,and R.S. Bucy,"New results in linear filtering and prediction theory",*J. Basic Engrg.*.**83**,95–108,1961.
[126] Kendall,M.G.,"On the analysis of oscillatory time-series", *J. Roy. Statist. Soc.,***A108,**93–109,1945.
[127] Kendall,M.G.,and A.Stuart,*The Advanced Theory of Statistics,* Vol.3,Charles Griffin,London,1966.
[128] Kitagawa,G.,and W.Gersch,"A smoothness priors-state space modeling of time series with trend and seasonality",*J. Amer. Statist. Assoc.,***79**,378–389,1984.

[129] Kohn,R.,and C.F. Ansley,"Estimation ,prediction,and interpolation for ARIMA models with missing data",*J. Amer. Statist. Assoc.*,**81**,751–761,1986.

[130] Kolmogoroff, A., "Sur l'interpolation et l' extrapolation des suites stationnaires",*C.R.Acad.Sci.Paris*,**208**,2043–2045,1939.

[131] Kolmogoroff,A.,"Stationary sequences in Hilbert space", *Bull. Math. Univ. Moscow*,**2**(6),1–40,1941.

[132] Kolmogoroff,A.,"Interpolation und Extrapolation von stationären zufälligen folgen", *Bull. Acad. Sci. (Nauk)U. S. S. R., Ser Math.*,**5**,3–14,1941.

[133] Koopmans, L. H., *The Spectral Analysis of Time Series,* Academic Press, New York,1974.

[134] Koopmans,T.,"Serial correlation and quadratic forms in normal variables", *Ann. Math. Statist.*,**13**,14–33,1942.

[135] Koopmans, T. C. (ed.) , *Statistical Inference in Dynamic Economic Models,* Wiley, New York, 1950.

[136] Kotnour, K. D. , G. E. P. Box, and R. J. Altpeter, "A discrete predictor-controller applied to sinusoidal perturbation adaptive optimization" , *Inst. Soc. Amer, Trans.* , **5**, 255–262, 1966.

[137] Ljung, G. M. , "Diagnostic testing of univariate time series models",*Biometrika,***73**,725–730,1986.

[138] Ljung,G.M.,and G.E.P.Box,"On a measure of lack of fit in time series models",*Biometrika,***65**,297–303,1978.

[139] Ljung,G.M.,and G.E.P.Box"The likelihood function of stationary autoregressive-moving average models",*Biometrika,***66**, 265–270,1979.

[140] Loève,M.,*Probability Theory I.*,Springer-Verlag,NewYork,1977.

[141] MacGregor,J.F.,"Topics in the control of linear processes with stochastic disturbances", Ph.D.thesis,University of Wisconsin-Madison,1972.

[142] Mann, H. B. , and A. Wald, "On the statistical treatment of linear stochastic difference equations", *Econometrica,* **11**, 173–220, 1943.

[143] Marquardt, D. W. , "An algorithm for least-squares estimation of nonlinear parameters", *J. Soc. Ind. Appl. Math.* , **11** 431–441, 1963.

[144] Martin,R.D.,and V.J.Yohai,"Influence functionals for time series"(with discussion),*Ann. Statist.*,**14**,781–855,1968.

[145] Montgomery,D.C.,and G.Weatherby,"Modeling and forecasting time series using transfer function and intervention methods", *AIIE Trans.*289–307,1980.

[146] Moran,P.A.P.,"Some experiments on the prediction of sunspot numbers",*J. Roy. Statist. Soc.*,**B16**,112–117,1954.

[147] Muth,J.F.,"Optimal properties of exponentially weighted forecasts",*J.Amer. Statist. Assoc.*,**55**,299–306,1960.

[148] Newbold, P. , "Bayesian estimation of Box-Jenkins transfer function-noise models", *J. Roy. Statist. Soc.* , **B35**, 323–336, 1973.

[149] Newbold,P.,"The exact likelihood function for a mixed auto-autoregressive-moving average process", *Biometrika,* **61**, 423–426, 1974.

[150] Newbold,P.,"The equivalence of two tests of time series model adequacy",*Biometrika*,**67**,463–465,1980.

[151] Noble, B. , *Methods Based on the Wiener–Hopf Technique for the Solution of partial Differential Equations,*Pergamon Press, Elmsford, N. Y. , 1958.

[152] Osborn,D.R.,"On the criteria functions used for the estimation of moving average processes",*J. Amer. Statist. Assoc.,* **77**,388–392,1982.

[153] Oughton,K.D.,"Digital computer controls paper machine",*Ind. Electron.,***3**,358–362,1965.

[154] Page,E.S.,"On problems in which a change in a parameter occurs at an unknown point",*Biometrika,***44**,248–252,1957.

[155] Page, E. S. , "Cumulative sum charts", *Technometrics,* **3**, 1–9, 1961.

[156] Pierce,D.A.,"Least squares estimation in dynamic- disturbance time series models",*Biometrika,***59**,73–78,1972.

[157] Pierce, D. A. , "Residual correlations and diagnostic checking in dynamic-disturbance time series models", *J. Amer. Statist. Assoc.* , **67**, 636–640, 1972.

[158] Plackett, R. L. , *Principles of Regression Analysis*, Clarendon Press, Oxford, 1960.

[159] Poskitt, D. S. , and A. R. Tremayne, "Testing the specification of a fitted autoregressive-moving average model", *Biometrika,* **67**, 359–363, 1980.

[160] Priestley, M. B. , *Non-linear and Non-stationary Time Series Analysis,* Academic Press, London, 1988.
[161] Quenouille,M.H.,"Approximate tests of correlation in time- series",*J. Roy. Statist. Soc.,***B11**,68–84,1949.
[162] Quenouille,M.H.,*Analysis of Multiple Time Series,* Hafner,New York,1957.
[163] Quenouille,M.H.,*Associated Measurements,*Butterworth, London,1952.
[164] Rao, C. R. , *Linear Statistical Inference and Its Applications,* Wiley, New York, 1965.
[165] Rao, J. N. K. , and G. Tintner, "On the variate difference method", *Aust. J. Statist.* , **5**, 106–116, 1963.
[166] Reinsel, G. , "Maximum likelihood estimation of stochastic linear difference equations with autoregressive moving average errors",*Econometrica,***47**,129–151,1979.
[167] Reinsel, G.C., and G.C.Tiao, "Impact of chlorofluoromethanes on stratospheric ozone:a statistical analysis of ozone data for trends" , *J. Amer. Statist. Assoc.* , **82**, 20–30, 1987.
[168] Reinsel,G.C.,and M.A.Wincek,"Asymptotic distribution of Parameter estimators for nonconsecutively observed time series", *Biometrika,***74**,115–124,1987.
[169] Rivera, D. E. , M. Morari, and S. Skogestad, "Internal model control. 4. PID controller design", *Ind. Eng. Chem. Process Des. Dev.* , **25**, 252–265, 1986.
[170] Roberts, S. W. , 'Control chart tests based on geometric moving averages", *Technometrics,* **1**, 239–250, 1959.
[171] Robinson, E. A. , *Multichannel Time Series Analysis,* Holden-Day, San Francisco, 1967.
[172] Said, S. E., and D. A. Dickey, "Hypothesis testing in ARIMA $(p, 1, q)$ models", *J. Amer. Statist. Assoc.* , **80**, 369–374, 1985.
[173] Savage, L. J. , *The Foundations of Statistical Inference,* Methuen, London, 1962.
[174] Schuster, A. , "On the investigation of hidden periodicities" , *Terr. Mag. Atmos, Elect.* ,**3**, 13–41,1898.
[175] Schuster,A.,"On the periodicities of sunspots",*Philos. Trans. Roy. Soc.,***A206**,69–100,1906.
[176] Schwarz, G. , "Estimating the dimension of a model", *Ann. Statist.* , **6**, 461–464, 1978.

[177] Shewhart,W.A.,*The Economic Control of the Quality of Manufactured Product*,Macmillan,New York,1931.

[178] *Short Term Forecasting*, ICI Monograph 2, Oliver & Boyd, Edinburgh, 1964.

[179] Silvey, S. D. , "The Lagrangian multiplier test", *Ann. Math. Statist.* , **30**, 389–407, 1959.

[180] Slutsky, E. , "The summation of random causes as the source of cyclic processes"(Russian), *Problems Econom. Conditions,* **3,** 1, 1927; English trans. , *Econometrica,* **5,** 105–146, 1937.

[181] Solo,V.,"The order of differencing in ARIMA models", *J. Amer. Statist. Assoc.,***79,**916–921,1984.

[182] Stokes,G.G.,"Note on searching for periodicities",*Proc. Roy. Soc.,***29**,122–123,1879.

[183] Stralkowski, C. M. , "Lower order autoregressive-moving average stochastic models and their use for the characterization of abrasive cutting tools", Ph. D. thesis. University of Wisconsin–Madison, 1968.

[184] Subba Rao, T. , and M. M. Gabr, *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Springer- Verlag, Berlin, 1984.

[185] Thompson,H.E. and G.C. Tiao,"Analysis of telephone data:a case study of forecasting seasonal time series",*Bell J. Econom. Manage. Sci.,***2**,515–541, 1971.

[186] Tiao.G.C.,G.E.P.Box,and W.J.Hamming,"Analysis of Los Angeles photochemical smog data:a statistical overview",*J. Air Pollut. Control Assoc.,***25**,260–268,1975.

[187] Tintner, G. , *The Variate Difference Method,* Principia Press, Bloomington, Ind. , 1940.

[188] Tong,H.,*Threshold Models in Non-linear Time Series Analysis,* Springer-Verlag,New York,1983.

[189] Tong, H. , and K. S. Lim, "Threshold autoregression,limit cycles and cyclical data", *J.Roy. Statist. Soc.* , **B42,** 245–292, 1980.

[190] Tasy,R.S.,"Time series model specification in the presence of outliers",*J.Amer. Statist. Assoc.,***81**,132–141,1986.

[191] Tsay,R.S.,"Identifying multivariate time series models", *J.Time Ser.Anal.,***10**,357–372,1989.

[192] Tsay, R. S. , and G. C. Tiao, "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models", *J. Amer. Statist. Assoc.* , **79**, 84–96, 1984.

[193] Tsay,R.S.,and G.C.Tiao,"Use of canonical analysis in time series model identification",*Biometrika*,**72**,299–315,1985.

[194] Tukey, J. W. "Discussion,emphasizing the connection between analysis of variance and spectrum analysis", *Technometrics,* **3**, 191–219, 1961.

[195] Walker,G.,"On periodicity in series of related terms", *Proc. Roy. Soc.*,**A131**,518–532,1931.

[196] Whittle,P.,"Estimation and information in stationary time series",*Ark. Math.*,**2**,423–434,1953.

[197] Whittle,P.,*Hypothesis Testing in Time Series Analysis* ,University of Uppsala,Sweden,1951.

[198] Whittle,P.,*Prediction and Regulation by Linear Least -Squares Methods,*English Universities Press,London,1963.

[199] Wichern, D. W. , "The behaviour of the sample autocorrelation function for an integrated moving average process",*Biometrika*, **60**, 235–239, 1973.

[200] Wiener,N.,*Extrapolation Interpolation and Smoothing of Stationary Time Series,*Wiley,New York,1949.

[201] Wilks,S.S.,*Mathematical Statistics,*Wiley,New York,1962.

[202] Wilson, G. , "Factorization of the covariance generating function of a pure moving average process", *SIAM J.Numer.Anal.*, **6**, 1–7,1969.

[203] Wilson, G. T. , "Optimal control:a general method of obtaining the feedback scheme which minimizes the output variance, subject to a constraint on the variability of the control variable", *Technical Report 20,* Department of Systems Engineering University of Lancaster, Lancaster, U. K. , 1970.

[204] Wilson,G.T.,Ph.D. thesis,University of Lancaster,U.K.,1970.

[205] Wincek, M. A. , and G, C. Reinsel, "An exact maximum likelihood estimation procedure for regression-ARMA time series models with possibly nonconsecutive data", *J. Roy. Statist. Soc.* , **B48**, 303–313, 1986.

[206] Winters, P. R. , "Forecasting sales by exponentially weighted moving averages", *Manage. Sci.* , **6**, 324–342, 1960.

[207] Wold, H. O. , *A Study in the Analysis of Stationary Time Scries,* Almqvist & Wiksell, Uppsala, Sweden, 1938(2nd. ed. 1954).

[208] Woodward,W.A.,and H.L.Gray,"On the relationship between the S array and the Box-Jenkins method of ARMA model identification", *J.Amer. Statist, Assoc.,***76**,579–587,1981.

[209] Yaglom, A. M. , "The correlation theory of processes whose nth difference constitute a stationary process",*Mat.Sb.*, **37**(79), 141, 1955.

[210] Yamamoto, T. , "Asymptotic mean square prediction error for an autoregressive model with estimated coefficients", *Appl. Statist.* , **35**, 123–127, 1976.

[211] Young, A. J. , *An Introduction to Process Control System Design*, Longmans, Green, New York, 1955.

[212] Yule, G. U. , "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers" , *Philos. Trans. Roy. Soc.* ,**A226**, 267–298, 1927.

[213] Zadeh, L. A. , and J. R. Ragazzini, "An extension of Wiener's theory of prediction",*J. Appl. Phys.,***21**,645,1950.

[214] Bray, J. , "Dynamic equations for economic forecasting with the G. D. P.-unemployment relation and the growth of G. D. P. in the United Kingdom as an example",*J. Roy. Statist. Soc.,* **A134,** 167–209, 1971.

[215] Hannan,E.J.,*Multiple Time Series,*Wiley,New York,1970.

[216] Haugh, L. D. , and G. E. P. Box, "Identification of dynamic regression(distributed lag)models connecting two time series", *J. Amer. Statist. Assoc.,* **72,** 121–130, 1977.

[217] Liu,L.M.,and D.M. Hanssens,"Identification of multiple- input transfer function models",*Comm. Statist.,***A11**,297–314,1982.

[218] Pankratz, A. , *Forecasting with Dynamic Regression Models,* Wiley, New York, 1991.

[219] Poskitt,D.S.,and A.R.Tremayne,"An approach to testing linear time series models",*Ann. Statist.,***9**,974–986,1981.

[220] Priestley,M.B.,*Spectral Analysis and Time Series,* Academic Press,New York,1981.

[221] Walker,A.M.,"Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series",*J. Aust. Math. Soc.,*4,363–384,1964.

中英人名索引

W

Y

Z

专题索引

A

D

E

J

K

N

Q

R